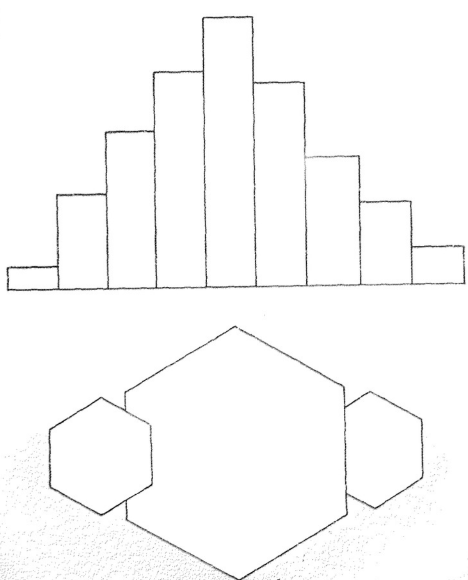
الاحتمالات والإحصاء

الدكتور محمد غالب المدني



مركز النشر العلمي
جامعة الملك عبد العزيز
جدة





الاحتمالات والإحصاء

الدكتور محمد غالب المدني
قسم الرياضيات الهندسية
كلية الهندسة - جامعة الملك عبد العزيز

مركز النشر العلمي
جامعة الملك عبد العزيز
ص ب ٨٠٢٠٠ - جدة ٢١٥٨٩
المملكة العربية السعودية

الطبعة الأولى :١٤١٢هـ (١٩٩٢م)

مطابع جامعة الملك عبد العزيز

تقــديم

تقوم طريقة البحث فى هذا الكتاب على الطريقة المتبعة فى البحوث العلمية ، فقد بدأتُ كتابي هذا بمقدمة بسيطة عرضت من خلالها الدور الهام الذى لعبته وتلعبه نظرية الاحتمال ، وكذلك فروع هذه النظرية . وضمَّنت كتابى أيضا ثمانية فصول وذلك تيسيرا للبحث ، وأوردت العديد من التمارين المحلولة وغير المحلولة ، والكثير من الرسومات التى تمكن الطالب من فهم وترسيخ المبادىء والتطبيقات الواردة فيه ، ولقد راعيت فى بعض الحالات سرد بعض النظريات ، وذكرت بعض العلاقات دون محاولة لبرهانها ، كذلك راعيت السهولة والبساطة فى انتقاء المواضيع الواردة فيه بدقة وبتسلسل لتُلائم ظروف وحاجات طلاب كلية الهندسة والعلوم ، وإمكانيات هذا المقرر .

ولست أدعى بحال أننى استوعبت تماما كل نواحى البحث والدراسة فى هذا الموضوع الهام ووصلت فيه إلى الغاية . فمجهود الفرد مهما عظم قليل ،وعلمه مهما خاض ضئيل « ربنا وسعت كل شيءٍ رحمة وعلماً » . ولابد من الاعتراف بأن هذا الكتاب جاء وليد مجهودات كثيرة ودراسات استمرت ست سنوات .

لذلك أرجو أن يكون هذا الكتاب فاتحة لبحوث جديدة ومنهجا صالحا لمحاولات صادقة نطمع أن ينهض بها علماؤنا .

أسأله تعالى أن يهدينا جميعا إلى صراطه المستقيم وألا يكلنا إلى أنفسنا وأن يوفقنا للوفاء بالعهد والإخلاص فى العمل لخير الإنسانية ، إنه نعم المولى ونعم النصير .

مقــدمة

تخضع جميع مظاهر حياتنا للسنن الكونية ، ولعل عجز العلماء عن التنبؤ بنتائج بعض هذه المظاهر من خلال قوانين علمية معروفة دفعهم للقول بأن هذه المظاهر تخضع لعامل الصدفة . فقد ينتهي بنا حادث غير مقصود إلى المستشفى ، ويدفعنا عجزنا عن السيطرة على الصدفة إلى اختيار الطريق الأفضل الذى يليها مباشرة . فنحاول أن نقدر احتمال حدوث حادثة معينة ينمّق كل منا بعبارات تدل على الصدفة . فنستعمل كلمات مثل « من المحتمل » ، « ربما » ، أو غيرها ، وعند تفكيرنا بحادثة لم تصبح حقيقة واقعة بعد ، أو كانت نتيجتها خارج نطاق سيطرتنا عليها ، فإننا نقوم عفويا بحساب الاحتمال والصدفة .

لقد نشأت نظرية المصادفة من خلال أبحاث بسكال وفيرما المتعلقة بألعاب الحظ المتنوعة . وتبدو فكرة خضوع الصدفة للقوانين غير مقنعة لمن يؤمن بسيطرة ما يسمى بالحظ ، ولكن قوانين الصدفة لا تنفى إمكانية فوز الإنسان فى بعض الأحيان بضربة حظ موفقة ، ولا تنكر أيضا قيمة الحدس فى التنبؤ ببعض الحوادث . ولا تصبح هذه القوانين بحق إلا عندما تكثر الأسئلة التى تدخل فى البحث ، مثل رميات متعددة لأحجار النرد ، وتوزيع أوراق اللعب عدداً كبيراً من المرات ، واصطدامات كثيرة بين السيارات ، وحساب أعمار عدد كبير من الناس ونشأة الحياة .

لقد استقت نظرية الاحتمال جذورها من المشتغلين بألعاب الحظ الذين كانوا يحاولون استشفاف معلومات تساعدهم على الفوز فى لعب الورق والنرد . ولم يبدأ حساب الاحتمال بالشكل الذى نعرفه اليوم إلا فى منتصف القرن السابع عشر على يد ثلاثة من الفرنسيين هم : دوفرما ، بسكال ، ودوميريه .

وقد برز فى هذه الفترة ، لدى الرياضيين عن أبحاثهم فى موضوع الاحتمالات ، اختصاص مستقل تماما عن أعمالهم هو « رياضيات الاحتمال » . وتطور هذا الاختصاص الجديد تطورا ملموسا حتى أصبح يسيطر على مظاهر عديدة من مظاهر الحياة الحديثة . فإلى جانب سيطرته على التأمين ساعد العالم الذرى فى فهم الآثار المتشابكة التى تسجلها الجزيئات الذرية المقذوفة من السيكلوترون على الأفلام ، وساعد خبير الصواريخ فى تحديد عوامل الأمان التى يجب أن تزود بها أجهزة القذائف الباهظة الثمن ، وساعد علماء النفس فى تقدير ذكاء الأطفال عند إجراء اختبارات الذكاء عليهم ، وساعد رجال الانتخابات فى توقع النتائج قبل حدوثها بيوم واحد ، كما ساعد عمال المصانع فى تدقيق السلع المنتجة وهى تتدحرج على خطوط الإنتاج فى المعمل .

وأصبح لنظريات الاحتمال والمصادفة من الأسس الرياضية السليمة ما يجعلها تطبق على نطاق واسع حيثما انعدم الحكم الصحيح المطلق ، وتقدمت دراسة نظريات المصادفة والاحتمال من الوجهة الرياضية تقدماً كبيراً ، حتى أصبحنا قادرين على التنبؤ بحدوث بعض الظواهر التى نقول إنها تحدث بالمصادفة ، والتى لا نستطيع أن نفسر ظهورها بطريقة أخرى (مثل قذف حجر النرد) . وأصبحنا بفضل هذه الدراسات قادرين على التمييز بين ما يمكن أن يحدث بطريقة المصادفة وما يستحيل حدوثه بهذه الطريقة .

هذا ويضم الاحتمال بين طياته قانونين هامين ، أولهما يسمى بقانون « أحد الحادثين » ويسمى الآخر بقانون « كلا الحادثين » . كما يضم أيضا قانونا هاما يعرف بقانون الأعداد الكبيرة .

وتكمن إحدى الصعوبات الكبيرة عند تطبيق قوانين الاحتمال فى تحديد الأساليب الممكنة عليها والتى يمكن أن تحدث بها حادثة ما . وليست هذه المشكلة صعبة جداً فى ألعاب الحظ . فقد تمكن علماء الاحتمال النظرى من إيجاد قوانين المتبادلات والمتوافقات والتى سهلت عليهم حل القضايا الصعبة ، بعد تفكيرهم وتأملهم فى الأنظمة والتراتيب التى يجري بها أى نوع من أنواع اللعب .

وبالرغم من أن حساب الاحتمال لايزال يحمل آثار اللعب واللهو اللذين اشتق منهما ، إلا أن ألعاب الورق والنرد ليست كل شيء فيه . فهو يشكل فى مظهره العملي العنصر الرئيس فى علم الإحصاء ويدخل حساب الاحتمال الإحصائى مجال الأعمال التجارية ، فيقدر كمية البضائع التى يجب على المنتج أن يخزنها فى مخازنه احتياطا ، ليستطيع تغطية

الطلبات غير متوقعة على منتوجاته ، كما يكشف لمهندسى المواصلات عدد الاتصالات التى ينبغى أن تنشأ فى أى مقسم هاتفى أو شبكة برقية . ويستخدم أيضا فى صناعة الأدوية ليكشف عما إذا كانت الآثار التى تنتج من استعمال متطوعين لدواء جديد هى آثار ذات قيمة إحصائية أم أنها مجرد صدفة . ويبقى فرق أساسى بين ألعاب الحظ وبين هذه التطبيقات الأخرى التى تفوق الأولى تعقيدا وتفضلها فائدة . وهو أنه يمكن دائما فى ألعاب الحظ أن نعدد كل النتائج الممكنة التى يمكن أن توزع بها أوراق اللعب . وقد يكون هذا التعداد صعبا إلا أنه ممكن دوما . أما عندما نرغب فى التنبؤ بتقلبات الحياة ، فكثيرا ما يستحيل أن نعرف مسبقا كل أوراق اللعب المطروحة .

والرياضى الذى يحسب احتمالات الحظ يسحب أوراقه من مجموعة يعرف مسبقا أرباحها ، كما يعرف قيمها النسبية . أما فى الاحتمال الإحصائى ، فإن محتويات الجعبة التى يتم السحب منها غير معروفة ، ولابد للإنسان أن ينتقى عينة تجريبية منها وأن يحسن انتقائها ، ثم يقدر احتمال كونها ممثلة لمحتويات الجعبة تمثيلا صادقا .

ويشبه حساب الاحتمال ، ومساعده الإحصاء ، شخصين يتجهان نحو منزل واحد من النهايتين المختلفتين لطريق واحد . ففى الاحتمال تعرف كل العوامل المؤثرة فى القضية . غير أن النتيجة يتكهن بها تكهنا . أما فى الإحصاء ، فإن النتائج معروفة ولكن العوامل التى تسببها مشكوك فى أمرها . وتوضح إحدى ألعاب حجر النرد حساب الاحتمال إيضاحا حسنا . فحجرا النرد يستطيعان إنتاج 36 توفيقة مختلفة ، ويبقى حساب احتمالاتها طوع يمين كل من يستطيع العد .

يمكن أن نخلص إلى القول بأن علم الاحتمال يبحث فى الظواهر العفوية والعلاقات القانونية التى تخضع لها . أما الإحصاء فهو يحوى ضمنا حساب الاحتمال ويحوى إلى جانبه أشياء أخرى تعتبر جزءا مما هو متعارف عليه علميا بالحساب الاحتمالى .

وما برح علم الاحتمال ينمو نمواً لا حدود له ، شأنه فى ذلك شأن بقية العلوم ، ولم يسبق أن رمى برج عاجى بظله الطويل على عالم الحياة اليومية ، مثلما فعل برج علم الاحتمال .

المحتويات

الفصل الأول

الاحتمـــال

■ فضاء العينة ■ الحوادث ■ العمليات على الحوادث ■ مبادىء العد ■ احتمال أى حادث ■ قوانين الاحتمال ■ الاحتمال الشرطى ■ نظرية بيز ■ تمارين محلولة ■ تمارين عامة .

(١,١) فضاء العينة Sample space

يحاول خبراء الإحصاء تفسير نتائج الحظ التى تقابل المشتغلين في الأبحاث العلمية . فهم يهتمون مثلا بعدد حوادث السير التي تقع شهرياً عند تقاطع شارعي (فلسطين — المدينة) فى مدينة جدة بالمملكة العربية السعودية محاولين تبرير تركيب إشارات ضوئية فى هذا التقاطع ، كما أنهم يهتمون بحجم الغاز المنطلق خلال تجربة كيميائية معينة وغير ذلك من الأمور .

إن المعلومات المسجلة بشكلها الخام كأعداد أو قياسات تدعى بالبيانات الخام (raw data) ، فمثلا تشكيل الأعداد 6 ,5 ,2 ,1 ,0 ,2 الممثلة لعد حوادث السير التي وقعت في الثلث الأول من هذا العام عند تقاطع شارعى (فلسطين — المدينة) بمدينة جدة مجموعة من البيانات الخام ، كذلك فإن مجموعة القياسات 17 ,15 ,25 بالسنتيمتر المكعب والممثلة لحجم الغاز المنطلق في عملية كيميائية معينة تمثل أيضا مجموعة من هذه البيانات .

ويستخدم الإحصائى كلمة تجربة ، أو تجربة إحصائية لوصف أى عملية تقدم له مجموعة بيانات خام . ومن أبسط الأمثلة على التجارب الإحصائية إلقاء قطعة نقود ، إطلاق قذيفة وملاحظة سرعتها فى فترات زمنية معينة ، كذلك معرفة الناخبين بالنسبة للمرشح X مثلا إلخ ..

تعريف (١,١) فضاء العينة

نسمى مجموعة كل النتائج الممكنة لتجربة إحصائية بفضاء العينة ونرمز لها بالرمز S .

إن كل نتيجة فى الفضاء S تدَعى عنصراً (element) أو نقطة عينة (sample-point) .

إذا احتوى فضاء العينة على عدد محدود من النقاط ، فإن باستطاعتنا ترتيب هذه العناصر ضمن قوسين يفصل بين أى منها فاصلة ، فمثلا إن فضاء العينة S الممثل لمجموعة نتائج

تجربة إلقاء حجر نرد يمكن كتابته بالشكل :

$$S = \{1, 2, 3, 4, 5, 6\}$$

أما إذا احتوى فضاء العينة على عدد غير محدود من النقاط ، فمن المتعذر عندئذ ذكر جميع عناصره ، وهنا نذكر الخواص التى تحققها عناصر هذا الفضاء ، فمثلا إذا كان S ممثلا لجميع مدن العالم والتى تعداد سكان أى منها أكثر من نصف مليون نسمة ، فإن S تكتب على النحو التالى :

S = { x | x مدينة عدد سكانها أكثر من نصف مليون نسمة }

كما أن مجموعة النقاط (x,y) الواقعة على محيط دائرة وداخلها ، نصف قطرها أربعة ومركزها المبدأ يمكن كتابتها على الشكل التالى :

$$S = \{(x, y) \mid X^2 + Y^2 \leq 16\}$$

إن طريقة تمثيل فضاء العينة بالشكل السابق له منفعة فى التجارب التي يكون فيها جدولة العناصر ممثلا لعمل شاق روتيني .

مثال (١,١)

لتكن تجربتنا قذف حجر نرد ، ولنفرض أن اهتمامنا ينصب على العدد الذى يظهر على حجر النرد لدى إلقائه ، فإن :

$$S_1 = \{1, 2, 3, 4, 5, 6\}$$

أما إذا كان اهتمامنا ينصب على نوع العدد زوجى أم فردى ، فعندئذ يكون فضاء العينة من الشكل :

S_2 = { زوجى ، فردى }

والمثال (١,١) يوضح لنا أن نتائج تجربة معينة يمكن تمثيلها بأكثر من فضاء عينة واحد / كما أن بعض هذه الأفضية قد تزودنا بمعلومات أكثر مما تزودنا بها بقية الأفضية . فمثلا نلاحظ أن S_1 يحدد لنا جميع عناصر الفضاء ، فإذا علمنا أن عنصراً من S_1 قد وقع فإن

باستطاعتنا أن نحدد أى النتائج في S_2 ستحدث ، ومع ذلك فإن معرفتنا بوقوع حادث فى S_2 لن يقدم لنا شيئا عن الحادث الذى وقع فى S_1 . وبشكل عام فإننا نستعمل فضاء العينة الذى يقدم لنا أكثر المعلومات حول نتائج التجربة المدروسة .

مثال (١,٢)

لنسحب أربعة عناصر بشكل عشوائى من مجموعة بضاعة مصنعة . ولنفحص هذه العناصر ، ولنصنف كلا منها بحسب نوعه معابا أو غير معاب . ففى الحالة الأولى نرمز له بالرمز D ، وفى الثانية بـ N . نلاحظ أن فضاء العينة الذى يزودنا بمعلومات أكثر هو :

$$S_1 = \{ NNNN, NNND, NNDN, NDNN, DNNN, NNDD, NDDN, NDND, DNND, DNDN, DDNN, NDDD\ DNDD, DDND, DDDN, DDDD \}$$

أما الفضاء $S_2 = \{ 0, 1, 2, 3, 4 \}$

فيوضح لنا عدد العناصر المعابة التى سحبناها ، وهو يزودنا بمعلومات أقل من S_1 .

(١,٢) الحوادث Events

سينصب اهتمامنا فى أى تجربة معطاة حول حدوث حادث معين أكثر من النتيجة الممثلة لعنصر ما فى فضاء العينة ، فمثلا إذا اعتبرنا فى المثال (١,٢) ممثلا على النحو التالى :

A = { عدد العناصر المعابة أكثر من واحد }

فهذا يعنى أننا نهتم بالمجموعة الجزئية :

$$A = \{ DDNN, DNDN, NNDD, DNND, NDND, NDDN, NDDD, DNDD, DDND, DDDN, DDDD \}$$

من فضاء العينة S_1 . وكذلك الأمر إذا كنا نهتم بالحادث A الذى يقع فيما لو كان الوجه الذى ظهر فى المثال (١,١) يقبل القسمة على 3 فهذا يعنى أيضا أن :

$$A = \{ 3 , 6 \}$$

سنخصص بالنسبة لكل حادث مجموعة من نقاط العينة ، وهذه المجموعة هى جزء من فضاء العينة .

تعريف (١,٢) الحادث The event

الحادث هو مجموعة جزئية من فضاء العينة .

مثال (١,٣)

بفرض أن $S = \{t|t > 0\}$ ، حيث يمثل t العمر السنوى لمادة مشعة ، فإذا كان $A = \{t|t < 2\}$ فإننا نلاحظ أن A يمثل مجموعة العناصر المشعة التى تتلاشى قبل نهاية السنة الثانية لتكوينها . هذا المثال يوضح لنا أن أية مجموعة جزئية يمكن أن تحدد لنا حادثاً عناصره نفس عناصر المجموعة الجزئية .

تعريف (١,٣) الحادث العينى The sample event

إذا حوى حادث ما على نقطة عينة واحدة من فضاء العينة ، قلنا عنه إنه حادث عينى ، أما الحادث المركب فهو الحادث الذى نعبر عنه من خلال اجتماع عدة حوادث عينية .

مثال (١,٤)

فى تجربة سحب ورقة من ورق اللعب فإن فضاء العينة :

S = { ديناری ، بستونى ، سباتى ، كبة }

أما الحادث { ديناری } = A فهو يمثل حادثا عينياً . أما الحادث C الممثل لسحب ورقة حمراء فهو حادث مركب لأن :

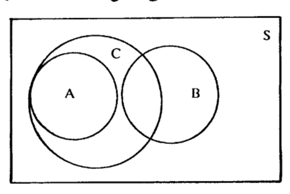
C = { ديناری ، سباتى }

تعريف (١,٤) الفضاء الخالى The empty space or the impossible event

إذا لم يحو فضاء العينة على أية نقطة عينة ، قلنا إنه فضاء صفرى أو فضاء خالى ونرمز بالرمز ϕ

لتكن التجربة سحب كرة سوداء من صندوق يحوى على كرات حمر فقط . نلاحظ أن فضاء العينة فى هذه التجربة هو الفضاء الخالى .

يمكن تمثيل العلاقة بين الحوادث وفضاء العينة بواسطة بما يسمى بمخطط فين ، وفى هذا المخطط نمثل فضاء العينة بمستطيل والحوادث المنبثقة عنه بدوائر فى داخله . ففى الشكل (١,١) نلاحظ أن الحوادث A,B,C تمثل جميعها مجموعات جزئية من فضاء العينة S . نلاحظ أيضا أن الحادث A هو مجموعة جزئية من الحادث C ، كما أن A,B لا يحويان نقاطا مشتركة ، وأن فى الحادثين C,B على الأقل نقطة عينة مشتركة .



الشكل (١,١)

(١,٣) العمليات على الحوادث Operations on events

يمكن أن نربط بين الحوادث لكى تكون حوادث جديدة ، وذلك باستعمال كافة أنواع العمليات المعروفة على المجموعات .

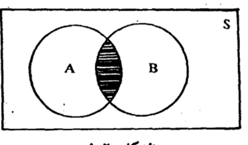
تعريف (١,٥) التقاطع The intersection

إن تقاطع حادثين B,A والذى نرمز له بالرمز $A \cap B$ أو A.B هو الحادث الذى يقع فيما لو وقع A و B معاً .

والعناصر المنتمية إلى $A \cap B$ يجب أن تكون موجودة فى كل من A و B في نفس الوقت . هذا ويمكن تمثيل التقاطع بواسطة طريقة القاعدة .

$$A \cap B = \{ X \mid x \in A,\ x \in B \}$$

يوضح القسم المظلل على الشكل (١,٢) عملية التقاطع .



الشكل (١,٢)

مثال (١,٥)

بفرض أن $A = \{1,2,3,4,\}$ ، $B = \{4,6,7\}$ عندئذ نجد أن $A \cap B = \{4\}$.

مثال (١,٦)

عند إلقاء قطعة نقود ثلاث مرات وملاحظة تتابع الصورة H والكتابة T فإننا نجد أن :

$$S = \{HHH, HHT, HTH, THH, THT, TTT, TTH, HTT\}$$

بفرض أن A هو الحادث الذى يقع فيما لو ظهر صورتين أو أكثر على التوالى ، وأن B هو الحادث الذى يقع فيما لو ظهرت نفس النتائج فى الرميات الثلاث عندئذ نجد أن :

$$A = \{HHH, HHT, THH\}$$

$$B = \{HHH, TTT\}$$

أما الحادث الممثل لتقاطع الحادثين السابقين فهو الحادث :

$$A \cap B = \{HHH\}$$

وهو الحادث المكون من ظهور الصورة فقط فى الرميات الثلاث . كما نلاحظ أن ظهور الصورة خمس مرات فى هذه التجربة هو المجموعة الخالية أو الحادث المستحيل .

مثال (١,٧)

فى المثال (١,١) وجدنا أن فضاء العينة :

$$S_1 = \{1,2,3,4,5,6\}$$

يتكون من ستة نتائج ممكنة للتجربة . فإذا فرضنا أن الحادث A يمثل ظهور وجه زوجى ، B يمثل ظهور وجه فردى ، وأن C يمثل ظهور عدد أولى عندئذ نجد أن :

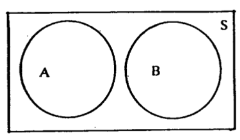
$$A = \{ 2,4,6 \}, \quad B = \{ 1,3,5 \}, \quad C = \{ 2,3,5 \}$$

وأن $A \cap B = \phi$ لأنه لا يمكن ظهور عدد زوجي وفردى بنفس الوقت .

تعريف (١,٦) الحادثين المتنافيين تبادلياً The utually exclusive events

نقول بأن الحادثين B , A متنافيين تبادلياً إذا كان $A \cap B = \phi$ ، وبمعنى آخر إذا تعذر وقوعهما المشترك .

يمكن توضيح مفهوم الحادثين المتنافيين باستخدام مخطط فين كما فى الشكل (١,٣) .



الشكل (١,٣)

نلاحظ فى المثال (١,٧) أن الحادثين B,A متنافيان .

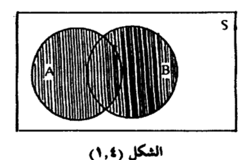
تعريف (١,٧) الاتحاد The union

إن اتحاد حادثين B,A والذى نرمز له بالرمز $A \cup B$ أو $A + B$ هو الحادث الذى يقع فيما لو وقع A أو B أو كلاهما معا .

هذا ويمكن توضيح عناصر الاتحاد بالطريقة التالية :

$$A \cup B = \{ x \mid x \epsilon A \text{ or } x \epsilon B \}$$

يوضح القسم المظلل على الشكل (١,٤) اتحاد الحادثين $A \cup B$



الشكل (١,٤)

ففي المثال (١,٧) نلاحظ أن $A \cup B = S$ وأن :

$$A \cup C = \{2,3,4,5,6\}$$

$$B \cup C = \{1,2,3,5\}$$

مثال (١,٨)

بفرض أن :

$$A = \{x \mid 0 < x < 3\}$$

$$B = \{y \mid 1 < y < 7\}$$

عندئذ نجد أن :

$$A \cap B = \{z \mid 1 < z < 3\}$$

$$A \cup B = \{z \mid 0 < z < 7\}$$

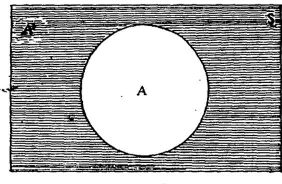
تعريف (١,٨) متمم حادث ما The complement of an event

إن متمم أي حادث A متعلق بالفضاء S هو مجموعة جميع العناصر من S غير الموجودة في A . هذا الحادث الجديد نرمز له بالرمز A'

نلاحظ أن :

$$A' = \{x \mid x \in S,\ x \in A\}$$

يمكن تمثيل المتمم بواسطة مخطط فين كما هو موضح على الشكل (١,٥) .



الشكل (١,٥)

نلاحظ في المثال (١,٧) أن متمم الحادث A هو الحادث B والعكس صحيح .

من التعاريف السابقة يمكن استنتاج النتائج الهامة التالية :

1. $A \cup \phi = A$
2. $A \cap \phi = \phi$
3. $A \cup A' = S$
4. $A \cap A' = \phi$
5. $(A')' = A$
6. $\phi' = S$
7. $S' = \phi$

(١,٤) مبادىء العد counting principles

سنورد فيما يلي بعض طرق تحديد عدد النواتج الممكنة لتجربة معينة بغير طريقة العد المباشر . وتسمى هذه الطرق باسم التحليل التوافقي .

القاعدة الأساسية للعد

نظرية (١,١)

إذا أمكن إجراء عملية معينة بعدد من الطرق المختلفة n_1 ، وإذا تلت هذه العملية عملية ثانية وأمكن إجراؤها بعدد n_2 من الطرق المختلفة ، فعندئذ يمكن إجراء العمليتين بعدد من الطرق مساو لـ $n_1 . n_2$

مثال (١,٩)

لدى إلقاء حجرى نرد دفعة ، واحدة ، فإن نقاط فضاء العينة لهذه التجربة هو 36 ، ذلك لأن الحجر الأول يمكن أن يعطى واحدا من ستة نتائج ، ومن أجل كل نتيجة من هذه النتائج ، فإن الحجر الثانى يمكن أن يعطى ستة نتائج أيضا . ولذلك حسب النظرية السابقة ، فإن عدد الثنائيات التى يقدمها الحجران معاً هو 36 . ويوضح الجدول (١,١) هذه النتائج كما يلي :

جدول (١,١)

النتائج					
(1, 1)	(1, 2)	(1, 3)	(1, 4)	(1, 5)	(1, 6)
(2, 1)	(2, 2)	(2, 3)	(2, 4)	(2, 5)	(2, 6)
(3, 1)	(3, 2)	(3, 3)	(3, 4)	(3, 5)	(3, 6)
(4, 1)	(4, 2)	(4, 3)	(4, 4)	(4, 5)	(4, 6)
(5, 1)	(5, 2)	(5, 3)	(5, 4)	(5, 5)	(5, 6)
(6, 1)	(6, 2)	(6, 3)	(6, 4)	(6, 5)	(6, 6)

هذا ويمكن تعميم النظرية (١,١) لتشمل أى عدد من الحوادث . والنظرية التالية توضح ذلك .

نظرية (١,٢)

إذا أمكن إجراء عملية ما بعدد n_1 من الطرق المختلفة ، وإذا تلت هذه العملية عملية ثانية وأمكن إجراؤها بعدد n_2 من الطرق المختلفة أيضا ، وإذا تلت هذه العملية الثانية عملية ثالثة وأمكن إجراؤها بعدد n_2 من الطرق المختلفة وهكذا ... عندئذ يكون عدد الطرق التى يمكن أن نجرى بها k عملية دفعة واحدة هو $n_1 . n_2 . n_3 ...... n_k$ طريقة .

مثال (١,١٠)

كم عدد فردى مؤلف من ثلاثة أرقام يمكن تشكيله من الأرقام 1,2,5,6,9 إذا استخدم كل رقم مرة واحدة ؟

الحل

بما أن العدد المطلوب فردى ، فلدينا إذاً ثلاثة خيارات من أجل كل رقم آحاد ، ومن أجل كل خيار من هذه الخيارات لدينا أربع خيارات لرقم العشرات وثلاثة خيارات لرقم المئات . وباستخدام النظرية (١,٢) نجد أن عدد الأعداد التى يمكن تشكيلها هو

$$3 \cdot 4 \cdot 3 = 36$$

فى أكثر الأحيان ، نهتم بفضاءات عينة تحتوى على عناصر من رتب مختلفة أو من فئات مرتبة من العناصر . فمثلا نريد أن نعرف عدد الترتيبات المختلفة الممكنة لجلوس ستة أشخاص على طاولة ، أو نريد أن نسأل السؤال التالى : ما هو عدد التشكيلات الممكنة لسحب ورقتين من مجموع 15 ورقة مرقمة ؟ إن هذه الترتيبات المختلفة تدعى تباديل (permutation)

تعريف (١,٩) التباديل Permutations

يسمى وضع n من الأشياء فى ترتيب معين بأنه تبديل لهذه الأشياء (بشرط أن تؤخذ جميع هذه الأشياء دفعة واحدة) كما يسمى وضع أى عدد $r \leq n, r$ من هذه الأشياء فى ترتيب معين بأنه تبديل العدد n من الأشياء مأخوذة r منها في كل مرة .

فمثلا إن تباديل أربعة حروف مأخوذة جميعها فى كل مرة هى :

abcd	bacd	cbad	dabc
abdc	badc	cbda	dacb
acbd	bcad	cdab	dbac
acdb	bcda	cdba	dbca
adbc	bdac	cabd	dcba
adcb	bdca	cadb	dcab

باستخدام النظرية (١,٢) نستطيع أن نصل إلى نفس النتيجة دون تعداد الحالات السابقة . فإذا تصورنا أربعة أمكنة فيمكن ملء المكان الأول بالأحرف الأربعة بأربع طرق ، أما المكان الثانى فيمكن ملؤه بثلاثة طرق ، والثالث بطريقتين ، والرابع بطريقة واحدة وهكذا يمكننا أن نملأ الأمكنة الأربعة دفعة واحدة بعدد من الطرق مساو لـ $(4)(3)(2)(1) = 24$ تبديلة . وبشكل عام فإن n من العناصر المختلفة يمكن ترتيبها بـ $n(n-1)(n-2) \ldots 3 \cdot 2 \cdot 1$ طريقة ، سنرمز للمضروب السابق بالرمز $n!$ ونقرؤه n عاملى . فمثلا يمكن ترتيب أربعة عناصر بـ $4! = 4.3.2.1$ طريقة . من التعريف السابق نجد أن $1! = 1$

نظرية (١,٣)

إن عدد تباديل n من الأشياء المختلفة هو $n!$. لنرى مثلا عدد التباديل الممكنة لحرفين من أربعة حروف a, b, c, d إن هذه التباديل هى :

ab,ac,ad,ba,ca,da,bc,cb,bd,db,cd,dc

ولشرح هذه الفكرة نتصور مكانين أو صندوقين . أما المكان الأول فيمكن ملؤه بأربعة حروف ، وبعد ملء المكان الأول بأحد الحروف الأربعة المفروضة ، فإنه يمكن ملء المكان الثانى بثلاثة طرق ، وحسب النظرية (١,٢) نجد أن عدد طرق ملء المكانين السابقين هو $4.3 = 12$ تبديلة . وبشكل عام نجد أن عدد تبديلات n شيئا مأخوذا منها r شيئا فى وقت واحد هو $n(n-1)(n-2)\ldots(n-r+1)$ طريقة .

سنرمز لهذا المضروب بالرمز

$$_nP_r = \frac{n!}{(n-r)!}$$

نظرية (١,٤)

إن عدد تباديل n شيئا مختلفة مأخوذة r شيئا منها فى نفس الوقت هو :

$$_nP_r = \frac{n!}{(n-r)!}$$

مثال (١,١٢)

ما هو عدد التباديل المكونة من ستة أشياء مأخوذة ثلاثة ثلاثة ؟

الحل

إن عدد التباديل هذه هو :

$$_nP_r = \frac{6!}{(6-3)!} = 120$$

وتعليل ذلك أنه لو تصورنا ثلاثة أمكنة فإنه يمكن شغل المكان الأول بستة طرق بوساطة هذه الأشياء الستة ، وبعد شغل المكان الأول بأحد هذه الأشياء ، يمكن شغل المكان الثانى بإحدى خمس طرق ، وبعد شغل المكانين الأول والثانى يمكن شغل المكان الأخير بأربع طرق وبحسب النظرية (١,٢) نجد أن هناك $6.5.4 = 120$ طريقة .

التباديل مع التكرار Permutations with repetition

يطلب فى بعض الأحيان معرفة عدد تباديل مجموعة من العناصر يكون بعضها متماثلاً فمثلا وجدنا أن عدد تباديل أربعة حروف a, b, c, d مختلفة مأخوذة جميعها دفعة واحدة هو 24 تبديلة مختلفة . فإذا فرضنا أن $a = b, c = d$ عندئذ يمكننا فقط ذكر التباديل التالية :

aacc, acac, caac, ccaa, acca, caca

وهذا العدد من التباديل المختلفة هو $6 = \frac{4!}{2!\,2!}$ تبديلة مختلفة .

نظرية (١,٥)

إن عدد التباديل المختلفة لـ n شيئا مأخوذة من k صنفاً بحيث تحتوى n_1 عنصرا من عناصر الصنف الأول ، n_2 من الثانى و .... n_k من الصنف الأخير هو

$$\frac{n!}{n_1!\,n_2!\,...\,n_k!}$$

مثال (١,١٢)

ما هو عدد الإشارات المختلفة التى يمكن تكوينها من بين ستة أعلام مرفوعة رأسياً بحيث يكون أربعة منها خضر واثنان بيض ؟

الحل

نلاحظ أنه توجد 15 إشارة مختلفة من أربعة أعلام خضر واثنان بيض .

$$\frac{6!}{4!\,2!} = 15$$

مثال (١,١٣)

ما هو عدد التباديل المختلفة التى يمكن تكوينها من جميع أحرف كلمة probability : نلاحظ أنه يوجد إحدى عشر حرفا منها اثنان متشابهان هما b وإثنان آخران متشابهان هما i وعدد التباديل المختلفة هو :

$$\frac{11!}{2!\,2!} = 9979200$$ تبديلة .

نهتم عادة بعدد الطرق التى يمكن فيها تقسيم n شيئا الى r مجموعة جزئية (تدعى كل واحدة منها خلية) وذلك باعتبار أن ترتيب هذه العناصر فى كل خلية غير هام . فمثلا إذا كان لدينا مجموعة الأحرف $\{x_1, x_2, x_3, x_4\}$ فإننا نلاحظ أن عدد الطرق التى يمكن بها تجزئة هذه المجموعة الى خليتين تحتوى الأولى منها على ثلاثة عناصر ، والثانية على عنصر واحد هو أربع طرق ، وهذه الطرق هى :

$$\{(x_1, x_2, x_3,), x_4\}, \quad \{(x_1, x_2, x_4), x_3\},$$

$$\{(x_1, x_3, x_4,), x_2\}, \quad \{(x_2, x_3, x_4), x_1\},$$

نرمز للعدد السابق بالرمز $4 = \frac{4!}{3!.1!} = \binom{4}{3,1}$ حيث يمثل العدد العلوى عدد العناصر أما العددين السفليين فيمثلا على الترتيب عدد عناصر الخليتين الأولى والثانية .

نظرية (١,٦)

إن عدد الطرق التى يتم بها تجزئة مجموعة مؤلفة من n شيئا مختلفا إلى r خلية ، تحتوى الأولى على n_1 شيئا ، الثانية n_2 شيئا ، وتحتوى الأخيرة على n_r شيئا هو

$$\binom{n}{n, n, \ldots, n_r} = \frac{n!}{n!n!\ldots,n_r!}$$

حيث إن $n_1 + n_2 + \ldots + n_r = n$

مثال (١,١٤)

ما هو عدد الطرق التى يمكن بها توزيع 12 قطعة نقدية مختلفة على ثلاثة أطفال ، بحيث يأخذ الطفل الصغير أربع قطع والأوسط خمس والكبير ثلاث قطع ؟

الحل

نلاحظ أن عدد التجزئات المرتبة لاثنتى عشرة قطعة نقدية إلى ثلاث خلايا تحتوى على الترتيب 3,5,4 هو

$$\frac{12!}{4!\,5!\,3!} = 27750$$

فى كثير من المسائل العملية نهتم بعدد الطرق التى يمكن أن نسحب بواسطتها r عنصرا من n عنصرا مفروضا دون الاهتمام بالترتيب . هذه السحبات تدعى بالتوافيق (combinations) ، والتوافيق هى بالضبط تجزئة مجموعة مؤلفة من n عنصراً مختلفا إلى خليتين تحتوى الأولى على r شيئا ، كما تحتوى الثانية على (n - r) شيئا المتبقية . سنرمز لهذا العدد من التوافيق بالرمز $\binom{n}{r}$

نظرية (١,٧)

إن عدد توافيق n شيئا متميزا مأخوذة منها r شيئا فى وقت واحد هو

$$\binom{n}{r} = \frac{n!}{r!(n-r)!}$$

مثال (١,١٥)

رجل له ثمانية عشر ولداً . ما هو عدد الطرق التى يمكنه فيها أن يدعو خمسة من أولاده إلى رفقته فى رحلة سيقوم بها إلى بلد أجنبى ؟

الحل

نلاحظ أن عدد الطرق التى يمكن أن يختار بها خمسة من 18 عنصرا مختلفا هو العدد

$$\binom{18}{5} = \frac{18!}{5!\,(18-5)!} = \frac{18!}{5!\,.\,13!} = 8568$$

مثال(١,١٦)

كم لجنة من ثلاثة أشخاص يمكن تشكيلها من خمسة أطفال وثلاث بنات بدون تحديد ؟

الحل

إن عدد اللجان هو

$$\binom{8}{3} = \frac{8\,!}{3\,!\,5\,!} = 56$$

(١,٥) احتمال أى حادث Probability of an event

الاحتمال هو أحد المفاهيم الأساسية الهامة فى الوقت الحاضر . نفرض أن كيساً يحتوى على ست كرات مختلفة الألوان متجانسة ومتماثلة ، ثلاث منها حمر واثنتان زرق ، وواحدة بيضاء ، مخلوطة خلطاً جيداً داخل الكيس . إن إمكانية سحب كرة ملونة بصورة عشوائية (أى سحب كرة حمراء أو زرقاء) هو أكبر من إمكانية سحب كرة بيضاء . والسؤال المطروح الآن : هل يمكننا وصف هذه الإمكانية بعدد ؟ من الواضح أن هذا ممكنا . إن العدد الذى يصف لنا هذه الإمكانية يدعى احتمالاً (probability) .

لنفرض أن فضاء العينة S هو فراغ منته ، سنرفق بكل نقطة عينة x_i من هذا الفضاء عددا نرمز له بالرمز w_i بحيث يكون $\Sigma w_i = 1$. فمثلا فى تجربة إلقاء حجر نرد نلاحظ أن جميع أوجه الحجر لها نفس إمكانية الوقوع ولذلك ، فإننا نرفق بكل وجه عدداً يساوى $\frac{1}{6}$ نسميه احتمال ظهور ذلك الوجه ، أما فى تجربة إلقاء قطعة نقود ، فإننا نلاحظ أن ظهور الصورة (H) له نفس إمكانية ظهور وجه الكتابة (T) ، ولذلك فإننا نرفق بكل وجه عددا مساويا لـ $\frac{1}{2}$ ، يمثل احتمال ظهور كل وجه . إن هذا النوع من فضاءات العينية يدعى بالفضاء ذى الاحتمالات المتساوية أو الفضاء المنتظم . وعلى وجه الخصوص إذا احتوى الفضاء المنتظم S على n من النقاط عندئذ نرفق بكل نقطة عينة من هذا الفضاء عدداً $\frac{1}{n}$ يكون ممثلا لاحتمال ظهور هذه النقطة .

لإيجاد احتمال وقوع حادث A نحسب مجموع كافة الأعداد المرفقة بنقاط هذا الحادث . نسمى هذا العدد بقياس الحادث A أو احتمال الحادث A ونرمز له بالرمز P(A) . وهكذا نجد أن قياس ϕ هو الصفر وقياس S هو الواحد .

تعريف (١,١٠) احتمال أى حادث Probability of an event

احتمال أى حادث A هو مجموع الأعداد المرفقة بمختلف نقاط العينة والموجودة فى هذا الحادث .

من هذا التعريف نستنتج النتائج الهامة التالية :

1. $\underline{P}(\phi) = 0$
2. $\underline{P}(S) = 1$
3. $0 \leq P(A) \leq 1$

مثال (١,١٧)

ألقيت قطعة نقود متوازنة ومتماثلة ثلاثة مرات متتالية . ما هو احتمال ظهور الصورة مرة واحدة على الأقل ؟

الحل

نلاحظ أن فضاء العينة S فى هذه التجربة هو :

$$S = \{ HHH, HTH, THH, HHT, TTH, THT, HTT, TTT \}$$

وبما أن القطعة متوازنة ، إذاً كل نتيجة ممكنة من هذه النتائج لها نفس إمكانية الوقوع . لذلك فإننا نفرق كل نقطة عينة بعدد w . وهكذا نجد أن $8w = 1$ أي إن $w = \frac{1}{8}$
لنرمز للحادث المطلوب بالرمز A ، عندئذ يكون :

$$P(A) = \frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8} = \frac{7}{8}$$

مثال (١,١٨)

صمم حجر نرد بحيث أن ظهور أى وجه زوجى له ضعف إمكانية ظهور أى وجه فردى . ما هو احتمال ظهور وجه أقل من 5 عند إلقاء هذا الحجر مرة واحدة ؟

الحل

نلاحظ أن $S = \{ 1,2,3,4,5,6 \}$
لنرفق بكل عدد فردى العدد w وبكل عدد زوجى من S العدد 2w بحيث يكون $9w = 1$
نلاحظ أن $w = \frac{1}{9}$ ، والاحتمال المطلوب للحادث A الممثل لظهور وجه أقل من 5 هو

$$P(A) = \frac{1}{9} + \frac{2}{9} + \frac{1}{9} + \frac{2}{9} = \frac{6}{9} = \frac{2}{3}$$

يمكن النظر إلى الأعداد المرفقة بنقاط أى حادث A على أساس أنها تمثل احتمالات الحوادث العينية التى يتألف منها الحادث A ، فإذا كانت التجربة من النوع الذى يمكن أن نفرض فيها أوزاناً متساوية لكل نقطة عينة من S ، فعندئذ يكون احتمال أى حادث وليكن A ممثلا للنسبة بين عدد عناصر الحادث A وعدد عناصر فضاء العينة S .

نظرية (١,٨)

إذا كان عدد إمكانيات تجربة معينة N إمكانية (متساوية فى إمكانية وقوعها) منها n إمكانية توافق الحادث A فعندئذ يكون احتمال هذا الحادث مساوياً .

$$P(A) = \frac{n}{N} : n \leq N$$

مثال (١,١٩)

اهو احتمال الحصول على ورقة دينارى عند سحب ورقة من ورق اللعب بصورة عشوائية ؟

الحل

إن عدد إمكانيات سحب ورقة من ورق اللعب هو $N = 52$.

ويوافق الحادث A الممثل لظهور ورقة دينارى عدداً من الحالات $n = 13$ ، وعليه فإن احتمال A هو

$$P(A) = \frac{n}{N} = \frac{13}{52} = \frac{1}{4}$$

مثال (١,٢٠)

ما هو احتمال ظهور وجه الصورة عند إلقاء قطعة نقود متوازنة ومتماثلة ؟

الحل

نلاحظ أن عدد النتائج الممكنة لدى إلقاء قطعة نقود متوازنة هو $N = 2$ ويوافق الصورة

من هذه النتائج $n = 1$ نتيجة ، وهكذا نجد أن احتمال الحادث B الممثل لظهور وجه الصورة هو

$$P(B) = \frac{n}{N} = \frac{1}{2}$$

مثال (١,٢١)

ما هو احتمال حصولنا على مجموع يساوى خمسة عند إلقاء حجرى نرد مرة واحدة ؟

الحل

بالعودة إلى المثال (١,١٠) نجد أن فضاء العينة يتألف من 36 نقطة عينة ويوافق الحادث A الممثل لظهور مجموع يساوى خمسة في الثنائيات التالية :

(2,3), (3,2), (4,1), (1,4)

وعددها أربع نتائج ممكنة . وهكذا نجد أن احتمال الحادث هو

$$P(A) = \frac{4}{36} = \frac{1}{9}$$

مثال (١,٢٢)

اخترنا نقطة بطريقة عشوائية من داخل دائرة . ما هو احتمال أن تكون هذه النقطة أقرب إلى مركز الدائرة منها إلى المحيط ؟

الحل

نفرض أن S مجموعة النقاط الواقعة داخل الدائرة ذات نصف القطر r ، وأن A مجموعة النقاط الواقعة داخل الدائرة المشتركة مع الدائرة الأولى فى نفس المركز ونصف قطرها $\frac{1}{3}$ r عندئذ

$$P(A) = \frac{\text{مساحة } A}{\text{مساحة } S} = \frac{M\left(\frac{1}{3}r\right)^2}{M r^2} = \frac{1}{9}$$

(١,٦) قوانين الاحتمال Probability Formulas

يبدو فى بعض الأحيان أن من السهل حساب احتمال حادث من الاحتمالات المعروفة لبعض الحوادث الأخرى . وهذا يبدو جليا إذا كان الحادث المراد حساب احتماله اجتماعا لحوادث أخرى معلومة الاحتمالات ، أو إذا كان متمما لحادث علم احتماله . سنذكر الآن بعض القوانين التى توفر لنا على الغالب حساب الاحتمال ، وأول هذه القوانين يدعى بقانون الجمع .

نظرية (١,٩)

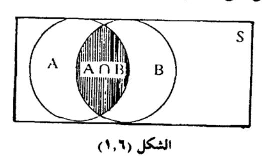
لأى حادثين A , B فإن

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

البرهان

بالعودة إلى مخطط فين الموضح بالشكل (١,٦) نجد أن $P(A \cup B)$ هو مجموع الأعداد المرفقة بنقاط العينة للحادث $A \cup B$. غير أن $P(A) + P(B)$ يمثل مجموع كل الأعداد المرفقة بنقاط العينة الموجودة فى الحادث A زائدا مجموع كل الأعداد المرفقة بنقاط العينة الموجودة فى B . لذلك فإننا نجمع الأعداد المرفقة بنقاط العينة الموجودة فى $A \cap B$ مرتين . وهذه الأعداد المرفقة بالتقاطع يجب طرحها من $P(A) + P(B)$ وهى تساوى .

$P(A \cap B)$ وبذلك نحصل على المطلوب .



الشكل (١,٦)

نتيجة (١,١)

اذا كان الحادثان B,A متنافيين فعندئذ يكون :

$$P(A \cup B) = P(A) + P(B)$$

إن النتيجة (١,١) تنتج مباشرة من النظرية (١,١٠) لأنه إذا كان B,A متنافيين فعندئذ يكون $A \cap B = \phi$ ، ومنه $P(A \cap B) = 0$. وبشكل عام يمكننا أن نكتب النتيجة التالية :

نتيجة (١,٢)

إذا كانت الحوادث $A_1, A_2, \ldots. A_n$ متنافية فعندئذ يكون :

$$P(A_1 \cup A_2 \cup \ldots. \cup A_n) = P(A_1) + P(A_2) + \ldots + P(A_n)$$

لاحظ أنه إذا كانت جملة الحوادث $A_1, A_2 \ldots. A_n$ مؤلفة لتجزئة لفضاء العينة S ، فعندئذ يكون :

$$P(A_1 \cup \ldots \cup A_n) = P(A_1) + \ldots + P(A_n) = P(S) = 1$$

مثال (١,٢٣)

بفرض احتمال أن يجتاز طالب امتحان الرياضيات هو 0.7 ، وامتحان الفيزياء هو 0.8 . فإذا علمت احتمال أن يجتاز واحد على الأقل من الامتحانين هو 0.9 ، فما هو احتمال أن يجتاز كلا الامتحانين ؟

الحل

إذا كان A ممثلا لاجتياز الطالب امتحان الرياضيات و B امتحان الفيزياء ، عندئذ نجد من النظرية (١,١٠) أن :

$$P(A \cup B) = P(A) + P(B) - P(A \cup B)$$

$$= 0.7 + 0.8 - 0.9 = 0.6$$

مثال (١,٢٤)

احسب احتمال الحصول على مجموع يساوى 5 أو 10 لدى إلقاء زوج من أحجار النرد .

الحل

لنفرض الحادث A الممثل لمجموع يساوى 5 ، والحادث B لمجموع يساوى 10 . نلاحظ

أنه يوافق الحادث A أربع نقاط عينية (من فضاء العينة المكون من 36 نقطة) انظر المثال (١,١٠) وهذه النقاط هى

$$(1,4),(4,1),(3,2),(2,3)$$

كما يوافق الحادث B ثلاث نقاط هى (6,4),(4,6),(5,5) ، ونظراً لكون هذه النقاط جميعها متساوية الإمكانية فإننا نجد أن $P(B) = \frac{1}{12}$ $P(A) = \frac{1}{9}$. كما نلاحظ أن الحادثين B,A متنافيان(لأنه لا يمكن الحصول على مجموع 10,5 بنفس الوقت فى نفس الإلقاء) لذلك نجد أن

$$P(A \cup B) = P(A) + P(B)$$

$$= \frac{1}{9} + \frac{1}{12}$$

$$= \frac{7}{36}$$

نظرية (١,١٠)

بما أن A, A' يمثلان حادثين متنافيين عندئذ يكون :

$$P(A') = 1 - P(A)$$

البرهان

نلاحظ أن $A \cup A' = S$ إذا

$$1 = P(S)$$

$$= P(A \cup A') = P(A) + P(A')$$

لذلك فإن :

$$P(A') = 1 - P(A)$$

مثال (١,٢٥)

ألقيت قطعة نقود متماثلة ومتوازنة أربع مرات متتالية . ما هو احتمال الحصول على وجه الصورة مرة على الأقل ؟

الحل

نفرض أن الحادث A يمثل ظهور وجه الصورة مرة على الأقل . نلاحظ أن فضاء العينة S يحتوى على عدد من نقاط العينة $16 = 2^4$ نقطة لانه يقابل كل إلقاء إحدى نتيجتين ، كما نلاحظ أن $P(A) = 1 - P(\acute{A})$ حيث أن $\acute{A}$ يمثل عدم ظهور أى صورة . ويمكن أن يقع $\acute{A}$ بشكل واحد فقط ولذلك فإن :

$$P(\acute{A}) = \frac{1}{16}$$

وأخيراً فإن

$$P(A) = 1 - \frac{1}{16} = \frac{15}{16}$$

(١,٧) الاحتمال الشرطى Conditional probability

إن احتمال وقوع حادث B عند معرفتنا بأن حادثا A قد وقع يدعى بالاحتمال الشرطى ونرمز له بالرمز $P(B|A)$ ، ونقرأ احتمال وقوع B علما أن الحادث A قد وقع ، أو بشكل أبسط احتمال B بمعلومية A .

لنفرض أن B يمثل ظهور عدد فردى عند إلقاء حجر نرد مرة واحدة . معتبرين أن حجر النرد مصمم بحيث تكون امكانية ظهور وجه فردى مساوية لضعف إمكانية ظهور وجه زوجى . فإذا أرفقنا بكل وجه زوجى العدد $\frac{1}{9}$ وبكل وجه فردى العدد $\frac{2}{9}$ فإننا نجد أن احتمال الحادث B هو $P(B) = \frac{2}{3}$ ، لنفرض الآن أننا نعلم أن نتيجة الإلقاء هى عدد أكبر من 2 . إذا نحن أمام فضاء عينة جديدة $A = \{3,4,5,6\}$. فلإيجاد احتمال وقوع الحادث B بالنسبة للفضاء الجديد A ، علينا أن نرفق أعدادا جديدة بعناصر هذا الفضاء بحيث يكون مجموع هذه الأعداد الواحد . فإذا أرفقنا بأى عدد زوجى w ، وبأى عدد فردى 2w ، فإننا نجد أن $6w = 1$ ، ومنه $w = 1/6$. وبالنسبة للفضاء A ، فإننا نجد أن B يحوى العنصرين 3,5 .

$$P(B/A) = \frac{4}{6} = \frac{2}{3}$$

هذا المثال يوضح أن لنفس الحادث B احتمالات مختلفة بالنسبة لفضاءات مختلفة . ويمكننا أيضا أن نكتب :

$$P(B/A) = \frac{2}{3} = \frac{4/9}{6/9} = \frac{P(A \cap B)}{P(A)}$$

حيث حسبنا $P(A), P(A \cap B)$ من فضاء العينة الأساس S . وبعبارة أخرى فإن الاحتمال الشرطي بالنسبة للفضاء الجزئي A من S يمكن حسابه مباشرة من الفضاء S نفسه .

تعريف (١,١١)

نعرف الاحتمال الشرطى للحادث B بمعلومية A ، والذى نرمز له بالرمز $P(B/A)$ بالعلاقة :

$$P(B/A) = \frac{P(A \cap B)}{P(A)}$$

شريطة أن يكون

$$P(A) > 0$$

مثال (١,٢٦)

ما هو احتمال الحصول على ثلاث صور لدى إلقاء ثلاث قطع نقود إذا علمت أنه قد ظهر صورة على القطعة الأولى لدى إلقائها ؟

الحل

نلاحظ أن فضاء العينة S يتألف من ثمانى نقط هى

$$S = \{HHH, TTT, HTT, THT, TTH, HHT, HTH, THH\}$$

فإذا كان على القطعة الأولى صورة ، فإننا نجد أن فضاء العينة الجزئى A هو

$$A = \{HHH, HTT, HHT, HTH\}$$

وحيث أن الحادث B الممثل لظهور ثلاث صور يوافقه حالة واحدة فى الفضاء A الذى يحتوى على أربع نقاط فاننا نجد أن

$$P(B/A) = \frac{P(A \cap B)}{P(A)} = \frac{1/8}{4/8} = \frac{1}{4}$$

الرمز $P(B/A)$ يمثل احتمال ظهور ثلاث صور بمعلومية أن القطعة الأولى قد أظهرت صورة .

مثال (١,٢٧)

عند إلقاء حجرى نرد متماثلين ومتوازيين ، وجد أن الوجهين الظاهرين مختلفان . ماهو احتمال :

١) أن يكون مجموع الوجهين اللذين ظهرا خمسة ؟
٢) أن يكون مجموع الوجهين أقل أو يساوى خمسة ؟

الحل

بالعودة إلى مثال (١,١٠) نلاحظ أن فضاء العينة S يحتوى على 36 نقطة عينة بينها ست نقاط يكون فيها الوجهان متماثلين وهذه النقاط هى :

$$(1,1),(2,2)(3,3),(4,4),(5,5),(6,6)$$

ولهذا فإن فضاء العينة المختزل A يحتوى على 30 = 6 - 36 نقطة عينة ، فإذا فرضنا أن الحادث B يمثل مجموع الوجهين الظاهرين خمسة فإننا نجد أنه يوافق هذا الحادث فى الفضاء المختزل أربع نقاط هى :

$$(2,3),(3,2),(4,1),(1,4)$$

وهكذا نجد أن الاحتمال الأول :

$$P(B) = \frac{4}{30} = 0.1333$$

أما إذا رمزنا للحادث الممثل لظهور مجموع يساوى خمسة أو أقل بالرمز C ، فإننا نجد أنه يوافق الحادث C النقاط التالية :

$$(1,2),(2,1),(1,3),(3,1),(1,4),(4,1),(2,3),(3,2)$$

ومنه :

$$P(C) = \frac{8}{30} = 0.2666$$

لاحظ أن الاحتمالات المحسوبة هى احتمالات شرطية بالنسبة لفضاء العينة غير المختزل . إذا ضربنا العلاقة الواردة فى التعريف (١,١١) بالاحتمال P (A) ، فإننا نستنتج نظرية الضرب التالية :

نظرية (١,١١)

إذا أمكن وقوع كلا الحادثين B,A فى تجربة معينة ، عندئذ نجد أن احتمال وقوعهما المشترك يعطى بالعلاقة التالية :

$$P(A \cap B) = P(A) \cdot P(B/A)$$

ولتوضيح طريقة استخدام النظرية (١,١١) نفرض أن لدينا علبة تحوى 15 فيوزا وبين هذا العدد من الفيوزات ثلاثة عاطلة (معابة) . لنسحب فيوزين بصورة عشوائية من العلبة على التتالى وبدون إعادة الفيوز الأول المسحوب إلى العلبة . ما هو احتمال أن يكون الفيوزان المسحوبان عاطلين (معابين) ؟ وللإجابة على هذا السؤال سنعتبر أن الحادث A يمثل { أن الفيوز الأول المسحوب عاطل } ، وأن B : { يمثل الفيوز الثانى المسحوب عاطل أيضا } . نلاحظ أن احتمال A هو $P(A) = \frac{1}{5}$ ، وأن $P(B/A) = \frac{1}{7}$ ولذلك فإن :

$$P(A \cap B) = P(A) \cdot P(B/A) = \frac{1}{5} \cdot \frac{1}{7} = 0.0526$$

ولتعميم النظرية (١,١١) نسوق النظرية التالية :

نظرية (١,١٢)

إذا أمكن فى تجربة معينة وقوع كل الحوادث $A_1, A_2, \ldots, A_n$ دفعة واحدة فعندئذ يكون احتمال الوقوع المشترك لهذه الحوادث جميعها معطى بالعلاقة التالية :

$$P(A_1 \cap A_2 \cap \ldots \cap A_n) = P(A_1)\, P(A_2/A_1)\, P(A_3/A_1 \cap A_2) \ldots$$

$$\ldots P(A_n/A_1 \cap A_2 \cap \ldots \cap A_{n-1})$$

ملاحظة

فى المثال السابق إذا أعيد الفيوز الأول المسحوب إلى العلبة فإننا نجد أن :

$$P(B/A) = P(B)$$

وهذا يعنى أن الحادثين B,A مستقلان .

تعريف (١,١٢) استقلال الحادثين Inpdendence of events

إن الشرط اللازم والكافى ليكون الحادثان B,A مستقلين هو أن يتحقق ما يلى :

$$P(A \cap B) = P(A).\ P(B)$$

مثال (١,٢٨)

لنحسب احتمال الحصول على مجموع يساوى 5 و 10 وذلك عند إلقاء حجرى نرد متوازنين ومتماثلين مرتين .

الحل

لنرمز للحوادث التالية :

$A_1 =$ { ظهور مجموع يساوى 5 على الوجهين فى القذفة الأولى }

$A_2 =$ { ظهور مجموع يساوى 5 على الوجهين فى القذفة الثانية }

$B_1 =$ { ظهور مجموع يساوى 10 على الوجهين فى القذفة الأولى }

$B_2 =$ { ظهور مجموع يساوى 10 على الوجهين فى القذفة الثانية }

نلاحظ أن الحوادث السابقة هى حوادث مستقلة . كما نلاحظ أن الاحتمال المطلوب هو :

$$P[(A_1 \cap B_2) \cup (B_1 \cap A_2)]$$

باستخدام النتيجة (١,١) نجد :

$$P[(A_1 \cap B_2) \cup (B_1 \cap A_2)] = P(A_1 \cap B_2) + P(B_1 \cap A_2)$$

وباستخدام التعريف (١,١٢) نجد أن :

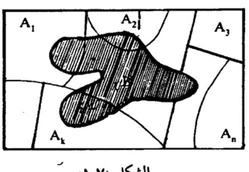
$$= P(A_1)\, P(B_2) + P(B_1)\, P(A_2)$$

$$= \left(\frac{4}{36}\right) \cdot \left(\frac{3}{36}\right) + \left(\frac{3}{36}\right) \cdot \left(\frac{4}{36}\right)$$

$$= 0.0185$$

(١,٨) نظرية بيز أو (قاعدة بيز) Bayes rule

لنفرض أن الحوادث $A_1, A_2, \ldots, A_n$ تشكل تجزئة منتهية لفضاء العينة S المتعلق بتجربة معينة (هذا يعني أن هذه الحوادث متنافية تبادليا . كما أن اتحادها يمثل الحادث الأكيد S) ، ولنفرض أنه يتعلق بنفس التجربة حادث آخر B احتماله $P(B) > 0$ كما هو موضح على الشكل (١,٧)



الشكل (١,٧)

عندئذ ستكون الحوادث $A_i \cap B$ متنافية تبادليا ، وذلك مهما يكن $i = 1, \ldots, n$ نلاحظ أن

$$B = S \cap B = [A_1 \cup A_2 \cup \ldots \cup A_n] \cap B$$

$$= (A_1 \cap B) \cup (A_2 \cap B) \cup \ldots \cup (A_n \cap B)$$

وباستخدام النظرية (١,١١) فإننا نجد أن :

$$P(B) = P(A_1) P(B/A_1) + P(A_2) P(B/A_2) + \ldots + P(A_n) P(B/A_n)$$

من ناحية أخرى نعلم أن :

$$P(A_i/B) = \frac{P(A_i \cap B)}{P(B)}$$

وبالتعويض نجد أن :

$$P(A_i/B) = \frac{P(B/A_i) \cdot P(A_i)}{P(A_1) P(B/A_1) + \ldots + P(A_n) P(B/A_n)} > 1 \leq i \leq n$$

نظرية (١,١٣) Bayes rule

لتكن $A_1, \ldots, A_n$ مجموعة من الحوادث المشكلة لتجزئة فضاء العينة S المتعلق بتجربة معينة ، وحيث أن $P(A_i) \neq 0$ من أجل جميع قيم $i = 1,\ldots,n$،لنفرض أن B حادثا يتعلق بنفس التجربة ويحقق الشرط $P(B) \neq 0$ ، عندئذ نجد أنه من أجل جميع قيم $i = 1,\ldots,n$

$$P(A_i/B) = \frac{P(B/A_i) \cdot P(A_i)}{\sum_{j=1}^{n} P(A_j)\, P(B/A_j)}$$

مثال (١,٢٩)

ثلاثة صناديق تحوى كرات حمراً ، وبيضاً ، وزرقاً كما هو موضح فى الجدول التالى :

الصندوق / اللون	الصندوق الأول	الصندوق الثانى	الصندوق الثالث
عدد الكرات الحمر	2	4	3
عدد الكرات البيض	3	1	4
عدد الكرات الزرق	5	3	3

سحبنا صندوقا بصورة عشوائية من بين الصناديق الثلاثة ، ثم سحبنا منه كرة بصورة عشوائية أيضا فكانت حمراء . ما هو احتمال أن يكون الصندوق المسحوب هو الأول ؟

الحل

نفرض الحوادث التالية :

A_i = { الصندوق المسحوب هو الصندوق ذو الرقم i وحيث إن $i = 1,2,3$ }

B = { الكرة المسحوبة من الصندوق حمراء }

فيكون المطلوب حساب $P(A_1/B)$

نلاحظ أن A_1, A_2, A_3 تمثل ثلاث حوادث متنافية تبادليا ، وأن اتحادها هو S (لأنه لابد من اختيار صندوق ، ولا يمكن أن نختار صندوقين دفعة واحدة ، كما أن

$$P(A_1) = P(A_2) = P(A_3) = \frac{1}{3}$$

كذلك فإن سحب كرة حمراء أى وقوع الحادث B يمكن أن يتم من الصندوق الأول ، أو الثانى ، أو الثالث أى إن B تقطع الحوادث A_1, A_2, A_3 ، كما أن $P(B) \neq 0$ وبحسب النظرية (١,١٣) نجد أن :

$$P(A_1/B) = \frac{P(B/A_1)\, P(A_1)}{P(B/A_1)\, P(A_1) + P(B/A_2)\, P(A_2) + P(B/A_3)\, P(A_3)}$$

لذلك فإن :

$$P(B/A_1) = \frac{2}{10}\ ,\ P(B/A_2) = \frac{4}{8}\ ,\ P(B/A_3) = \frac{3}{10}$$

وبالتعويض فى العلاقة السابقة نجد أن

$$P\,(A_1/B) = \frac{\frac{2}{10} \cdot \frac{1}{3}}{\frac{2}{10} \cdot \frac{1}{3} + \frac{4}{8} \cdot \frac{1}{3} + \frac{3}{10} \cdot \frac{1}{3}} = \frac{2}{10} = 0.2$$

تمارين محلولة

تمرين (١)

كم طريقة يمكن أن يجلس بها ثمانية أشخاص على مقعد يتسع لثلاثة أشخاص فقط ؟

الحل

لنتصور الأمكنة الثلاثة a, b, c كما هو موضح ☐ ☐ ☐ نلاحظ أن المكان a يمكن شغله بثماني طرق من قبل كل شخص . وفى حالة ملء المكان a بأحد الأشخاص الثمانية ، يبقى لدينا سبعة أشخاص وعندئذ يمكن اشغال المكان b بسبع طرق مختلفة ، وبعد شغل المكانين الأول والثانى بشخصين يبقى لدينا ستة أشخاص وست طرق لإشغال المكان c ، وعليه فإن عدد الطرق المطلوبة هو :

$$8.7.6 = 336$$

تمرين (٢)

كم عددا مؤلفا من أربعة أرقام يمكن تكوينها من الأرقام العشرة التالية 0, 1, 2, ... ,9 وذلك فى الحالتين التاليتين ؟

١ ـــ التكرار ممكن

٢ ـــ التكرار غير ممكن

الحل

الحالة الأولى : إذا كان تكرار أى رقم مسموح به فى عملية التشكيل ، فى هذه الحالة يمكن أن يكون الرقم الأول أى رقم من 9 ... ,1, 2 أما الأرقام الثانى والثالث والرابع

فيمكن أن يكون كلاً منها أحد الأرقام العشرة المفروضة . فيكون مجموع الأعداد المطلوبة هو

$$9.10.10.10 = 9000$$

عدداً .

الحالة الثانية : أما إذا كان تكرار أى رقم غير ممكن ، فإننا نلاحظ أن العدد المطلوب من الأعداد هو $9.9.8.7 = 4536$ عدداً .

تمرين (٣)

سحبنا ثلاث بطاقات من مجموع خمسين بطاقة . فإذا فرضنا أنه لا أهمية لترتيب عملية السحب ، فما هو عدد نقاط فضاء العينة فى هذه التجربة ؟

الحل

بحسب النظرية (١,٤) فإن فضاء العينة S يحتوى على عدد من النقاط مساوٍ لـ

$$\binom{50}{3} = \frac{50\,!}{(50-3)\,!\times 3\,!} = \frac{50!}{47!\times 3!} = \frac{(50)\,(49)\,(48)}{3\times 2\times 1} = 19600$$

تمرين (٤)

نود شراء آلة لصنع الحرير من إحدى خمس شركات . فبكم طريقة يمكن اختيار ثلاث من هذه الشركات ؟

الحل

حسب النظرية (١,٧) ، نجد أن عدد الطرق التى يمكن أن نختار بها ثلاث شركات من الخمس المفروضة هو :

$$\binom{5}{3} = \frac{5\,!}{3\,!\,2\,!} = \frac{5\,.\,4}{2} = 10$$

تمرين (٥)

إذا علمت $A \subset B$. فبين أن $P(A) \leq P(B)$

الحل

من الواضح أن :

$$P(B) = P(B \cap S) = P[B \cap (A \cup A')]$$

$$= P[(B \cap A) \cup (B \cap A')]$$

نلاحظ أن الحادثين $B \cap A$ ، $B \cap A'$ متنافيان ، لأنهما جزئين من الحادثين المتنافيين A' ، A . وبحسب النتيجة (١،١) نجد أن :

$$P(B) = P(B \cap A) + P(B \cap A')$$

وباعتبار أن $B \supset A$ إذن $B \cap A = A$ وهكذا فإن

$$P(B) = P(A) + P(B \cap A')$$

وبما أن $P(B \cap A') \geq 0$ إذن $P(B) \geq P(A)$

تمرين (٦)

بفرض أن احتمال سحب بطاقة رقمها مؤلف من ستة أعداد مجموع أول ثلاثة منها يساوى مجموع الثلاثة الأخر هو $P = 0.05525$. ابحث عن احتمال الحصول على بطاقة من هذا النوع ، وذلك من بين بطاقتين إذا كانت :

١) كلا البطاقتين لهما أرقام متتالية .

٢) إحدى البطاقتين مسحوبة بصورة عشوائية مستقلة عن البطاقة الثانية .

الحل

بفرض { البطاقة المسحوبة الأولى لها مجموعات متساوية } = A

{ البطاقة المسحوبة الثانية لها مجموعات متساوية } = B

أولا :

$$P(A \cup B) = P(A) + P(B) = 2.\ P(A) = 2.\ P = 0.1105$$

ثانيا : باستخدام النظرية (١،٩) والتعريف (١،١٢) نجد أن :

$$P(A \cup B) = P(A) + P(B) - P(A).\ P(B) = 2P - P^2 = 0.10744$$

تمرين (٧)

بفرض أن الحادث A يمثل ظهور وجه الصورة ، وأن الحادث B يمثل ظهور أحد الوجهين 2 أو 4 وذلك لدى إلقاء حجر نرد وقطعة نقود دفعة واحدة . أوجد :

١ — $P(A \cap B')$

٢ — $P(A/B)$

٣ — $P(A' \cup B')$

الحل

أولا : نلاحظ أن الحادث $A \cap B'$ يقع فيما لو ظهرت صورة على قطعة النقود ، ولم يظهر أحد الوجهين 2 أو 4 على حجر النرد لدى إلقائه وأن الحادثين A . B' مستقلان .

ويمكن بحسب النتيجة (١,١) والنظرية (١,١٠) ، أن نكتب :

$$P(A \cap B') = P(A) \,.\, P(B') = P(A)\,[\,1 - P(B)\,]$$

$$= \frac{1}{2}[\,1 - (\frac{1}{6} + \frac{1}{6})\,]$$

ومنه :

$$P(A \cap B') = \frac{1}{3}$$

ثانيا : أما الحادث (A/B) فيمثل ظهور A بمعلومية B ، وبما أن B,A مستقلين فإن :

$$P(A/B) = P(A) = \frac{1}{2}$$

ثالثا : وأخيرا فإن الحادث $A' \cup B'$ يمثل ظهور الكتابة على قطعة النقد ، أو أحد الأعداد 1,3,5,6 أو كلاهما معا . ونعلم حسب النظرية (١,٩) أن :

$$P(A' \cup B') = P(A') + P(B') - P(A' \cap B')$$

ثم باستخدام النظرية (١,١٠) نجد أن :

$$P(A') = 1 - P(A) = \frac{1}{2},\ P(B') = \frac{4}{6} = \frac{2}{3}$$

ومنه :

$$P(\acute{A} \cup \acute{B}) = \frac{1}{2} + \frac{2}{3} + \frac{1}{2} + \frac{2}{3}$$
$$= 0.8333$$

تمرين (٨)

إذا علمت أن الحادثين A,B متنافيان $(A \cap B = \phi)$ ، وأن احتماليهما غير معدومين ، فهل يكون B,A مستقلين ؟

الحل

من التنافى نجد أنه وقع A فلا يمكن أن يقع B ، وهذا يعنى أن $P(B/A) = 0$ ، وبما أننا فرضنا أن $0 < P(B)$ اذا $P(B)\ \ P(B/A)$ وهذا يعنى أن :

$$P(A \cap B) = P(A)\,P(B/A) \neq P(A)\,P(B)$$

وحسب النظرية (١,١١) نجد أن الحادثين B, A غير مستقلين .

تمرين (٩)

عند سحب ورقتين من ورق اللعب ما هو احتمال الحصول على ملكة وتسعة ؟

الحل

لنرمز للحوادث C,B,A كالتالى :

A = { الورقتان المسحوبتان ملكة وتسعة }

B = { الورقة الأولى ملكة والثانية تسعة }

C = { الورقة الأولى تسعة والثانية ملكة }

نلاحظ أن $A = B \cup C$ ، وأن C,B حادثين متنافيين ، لحساب P(B) ، P(C) نعرف الحوادث الآتية :

B_1 = { الورقة المسحوبة الأولى ملكة }

B_2 = { الورقة المسحوبة الثانية ملكة }

$B_3 =$ { الورقة المسحوبة الأولى تسعة }

$B_4 =$ { الورقة المسحوبة الثانية تسعة }

واعتماداً على التعريف (١,١١) نجد أن :

$$P(B) = P(B_1 \cap B_4) = P(B_1)\,P(B_4/B_1) = \frac{4}{52} \cdot \frac{4}{51}$$

$$P(C) = P(B_3 \cap B_4) = P(B_3)\,P(B_2/B_3) = \frac{4}{52} \cdot \frac{4}{51}$$

وحسب النتيجة (١,١) نجد :

$$P(A) = P(B) + P(C) = \frac{8}{663} = 0.01206$$

تمرين (١٠)

كتبنا على بطاقات خاصة الأحرف المكونة لكلمة عندان ، والبالغ عددها خمس بطاقات ، ثم خلطناها خلطا جيدا ، ثم سحبنا هذه البطاقات الواحدة تلو الأخرى . ما هو احتمال أن تشكل الأحرف المسحوبة كلمة عدنان ؟

الحل

نلاحظ أن كلمة عندان تتألف من أربعة حروف هي العين والنون والدال والألف وقد ورد حرف العين مرة واحدة ، والنون مرتين والدال مرة ، والألف مرة أيضا ، لذلك فإن :

$$P = \left(\frac{1}{5}\right) . \left(\frac{1}{4}\right) \left(\frac{2}{3}\right) . \left(\frac{1}{2}\right) . \left(\frac{1}{1}\right) = \frac{1}{60} = 0.0167$$

تمرين (١١)

احسب $P(A \cap B)$, $P(A \cup B)$ إذا علمت أن $P(A) = P(B) = \frac{1}{2}$ وأن $P(A/B) = \frac{2}{3}$

الحل

نلاحظ من النظرية (١,١١) أن :

$$P(A \cap B) = P(B). P(A/B) = \frac{1}{2} + \frac{2}{3} = \frac{1}{3}$$

وحسب النظرية (١,٩) نجد أيضا أن :

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \frac{1}{2} + \frac{1}{2} - \frac{1}{3} = \frac{2}{3}$$

تمرين (١٢)

راميان يصيب أحدهما الهدف باحتمال قدره 0.8 ويصيب الآخر نفس الهدف باحتمال قدره 0.7 . ماهو احتمال إصابة الهدف إذا أطلق منهما طلقة من بندقيته ؟

الحل

لنرمز بـ

$A =$ { أصاب الرامي الأول الهدف }

$B =$ { أصاب الرامى الثانى الهدف }

باستخدام النظرية (١,٩) نجد أن :

$$P(A \cup B) = P(A) + P(B) - P(A \cup B)$$

ومن الفرض لدينا :

$$P(A) = 0.8. P(B) = 0.7$$

وحسب التعريف (١,١٢) نجد أن :

$$P(A \cap B) = P(A). P(B) = (0.8). (0.7)$$

وهكذا نجد أن :

$$P(A \cup B) = 0.8 + 0.7 - (0.8) (0.7) = 0.94$$

تمرين (١٣)

رجل وزوجته عمرهما على الترتيب 45,60 عاما . واحتمال وفاة كل منهما خلال الأعوام العشرة القادمة هو على الترتيب 0.45, 0.7 فإذا علمنا أن أحدهما قد توفى ، فما هو احتمال أن يكون الرجل ؟

الحل

إذا رمزنا للحادثين B,A بالرمزين :

A = { وفاة الرجل خلال الفترة المحددة }

B = { وفاة أحد الشخصين }

فيكون المطلوب حساب P(A/B) . وحسب النظرية (١,١١) نجد أن :

$$P(A/B) = \frac{P(A \cap B)}{P(B)}$$

غير أن الحادث $A \cap B$ يمثل وفاة أحد الزوجين ووفاة الرجل بنفس الوقت ، أى بوفاة الرجل وبقاء الزوجة على قيد الحياة . وهذا الاحتمال يساوى حسب النظرية (١,١٠)

$$(0.7)[1-0.45] = 0.385$$

ثم أن :

$$P(B) = (0.7) + (0.45) - 2(0.7)(0.45) = 0.52$$

وبالعودة إلى العلاقة (★) نجد أن :

$$P(A/B) = \frac{0.385}{0.52} = 0.74038$$

تمرين (١٤)

كتب عدنان ثلاث رسائل إلى أصدقائه فى الولايات المتحدة ، ووضع كلا منها فى ظرف . واختلطت الظروف الثلاثة قبل أن يعنونها ، بعد ذلك عنون هذه الرسائل بصورة عشوائية . ما هو احتمال أن يكون قد كتب على ظرف واحد على الأقل العنوان الصحيح ؟

الحل

بفرض أن $A_k =$ { كتب عدنان العنوان الصحيح على الظرف ذو الرقم K }

حيث إن $K = 1,2,3$ ، فيكون المطلوب عندئذ حساب $P(A_1 \cup A_2 \cup A_3)$ باستخدام النظرية (١,٩) من أجل ثلاث حوادث نجد أن :

$$P(A_1 \cup A_2 \cup A_3) = [P(A_1) + P(A_2) + P(A_3)] - [P(A_1 \cap A_2) + P(A_1 \cap A_3) + P(A_2 \cap A_3)] + P(A_1 \cap A_2 \cap A_3)$$

غير أن :

$$P(A_k) = \frac{1}{3} : K = 1,2,3$$

$$P(A_1 \cap A_2) = P(A_1) . P(A_2/A_1) = \frac{1}{3} \cap \frac{1}{2} = \frac{1}{6}$$

$$P(A_1 \cap A_2) = \frac{1}{6} , P(A_2 \cap A_3) = \frac{1}{6}$$

وباستخدام النظرية (١,١٢) نجد أن :

$$P(A_1 \cap A_2 \cap A_3) = P(A_1) . P(A_2/A_1) . P(A_3/A_1 \cap A_3)$$

$$\frac{1}{3} . \frac{1}{2} . \frac{1}{1} = \frac{1}{6}$$

أخيرا وبالتعويض فى العلاقة (★) نستنتج الاحتمال المطلوب :

$$P(A_1 \cup A_2 \cup A_3) = [\frac{1}{3} + \frac{1}{3} + \frac{1}{3}] - [\frac{1}{6} + \frac{1}{6} + \frac{1}{6}] + \frac{1}{6}$$

$$= \frac{2}{3} = 0.666$$

تمرين (١٥)

لعب أحمد وسعيد عشر مباريات فى كرة الطاولة . كسب أحمد خلالها 4 مباريات ، كما كسب سعيد 3 مباريات وتعادلا فى 3 مباريات واتفقا على اللعب ثلاث مباريات أخرى ، ما هو :

١) احتمال أن يربح أحمد جميع المباريات ؟

٢) احتمال أن يكسب أحمد وسعيد بالتبادل ؟

٣) احتمال أن تنتهى مباراتان بالتعادل ؟

الحل

لنسمى الحوادث :

A = { كسب أحمد فى مباراة }

B = { كسب سعيد فى مباراة }

C = { تعادل أحمد وسعيد فى مباراة }

من شروط اللعب الأول نجد أن :

$$P(A) = \frac{4}{10} , P(B) = \frac{3}{10} , P(C) = \frac{3}{10}$$

أولا : احتمال أن يربح أحمد المباريات الثلاث يساوى احتمال أن يربح المباراة الأولى والثانية والثالثة .

$$P(AAA) = P(A) . P(A) . P(A) = (\frac{4}{10})^3 = 0.064$$

ثانيا : (إذا كسب أحمد ، ثم كسب سعيد ، ثم كسب أحمد) أو (كسب سعيد ، ثم كسب أحمد ، ثم كسب سعيد) فهذا يعنى أنهما تعادلا بالتبادل وهذا الاحتمال يحسب بالشكل التالى :

$$P [A \cap B \cap A) \cup (B \cap A \cap B)] = P (A \cap B \cap A) + P (B \cap A \cap B) =$$

$$P (A) . P (B) . P (A) + P (B) . P (A) . P (B) = 0.84$$

وبالتعويض نجد أن الاحتمال = 0.84

ثالثا : لحساب هذا الاحتمال نلاحظ أن إنهاء مباراتين بالتعادل يعنى انتهاء المباراتين الأولى والثانية ، أو الأولى والثالثة ، أو الثانية والثالثة بالتعادل ، وهذا يعنى أن الاحتمال المطلوب هو :

$$P[(C \cap C \cap \overline{C}) \cup (C \cap \overline{C} \cap C) \cup (\overline{C} \cap C \cap C)] =$$

$$3P (C) . P(C) . P \overline{C}) = 3. \frac{3}{10} . \frac{3}{10} . \frac{7}{10} = 0.189.$$

تمرين (١٦)

بفرض أن إنتاج ثلاثة أنواع من آلات حياكة الثياب هى على التوالى : 0.2,0.3,0.5 من الإنتاج الكلى لمصنع نسيج ، فإذا كانت نسبة الإنتاج المعاب لهذه الآلات هى على التوالى : 0.05, 0.04, 0.03 ، وإذا اخترنا قطعة بصورة عشوائية من الإنتاج العام ، ووجدنا أنها ذات عيب ، فما هو احتمال أن تكون هذه القطعة من حياكة الآلة ذات الإنتاج 0.5 ؟

الحل

لنرمز للآلات الثلاث بالرموز C,B,A على الترتيب ، وبالرمز D للقطعة المسحوبة إن كانت معابة ، فيكون المطلوب حسب النظرية (١,١٣) :

$$P(A/D) = \frac{P(D/A) \, . \, P(A)}{P(D/A) \, . \, P(A) + P(D/B) \, P(B) + P(D/C) \, . \, P(C)}$$

وهكذا نجد أن :

$$P(A/D) = \frac{(0.03)\,(0.5)}{(0.03)\,(0.5) + (0.04)\,(0.3) + (0.05)\,(0.2)}$$

تمرين (١٧)

مجموعتان من البضائع المصنعة ، عناصر المجموعة الأولى جميعا من النوع الجيد ، أما عناصر المجموعة الثانية فربعها معاب . اخترنا مجموعة بصورة عشوائية من بين المجموعتين ، ثم سحبنا عنصراً منها فظهر أنه جيد ، ما هو احتمال أن نسحب عنصرا آخر من المجموعة المسحوبة فيظهر أنه ذو عيب ، إذا علمت أننا أعدنا العنصر المسحوب فى المرة الأولى إلى نفس المجموعة المسحوبة ؟

الحل

لنسمى الحوادث التالية :

$A_1 =$ { المجموعة المختارة هى المجموعة الثانية }

$A_2 =$ { المجموعة المختارة هى المجموعة الأولى }

$B_1 =$ { العنصر المسحوب الأول من النوع الجيد }

من شروط المسألة نجد أن :

$$P(A_1) = P(A_2) = \frac{1}{2}$$

$$P(B/A_1) = \frac{3}{4} \, , \, P(B/A_2) = 1$$

وحسب العلاقة :

$$P(B) = P(B/A_1) \, . \, P(A_1) + P(B/A_2) \, . \, P(A_2)$$

نجد أن :

$$P(B) = \frac{3}{4} \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} = 0.875$$

وبعد أول سحب فإن احتمال أن تحوى المجموعة المختارة عنصراً غير جيد (معاب) يحسب من النظرية (١,١٣) :

$$P(A_1/B) = \frac{P(A_1) . P(B/A_1)}{P(A)}$$

$$P(A_1/B) = \frac{\frac{1}{2} . \frac{3}{4}}{0.875} = \frac{3}{7} = 0.42857$$

أما احتمال أن تحتوى المجموعة المختارة على عنصر جيد فهو :

$$P(A_2/B) = \frac{4}{7}$$

لنرمز للحادث | العنصر الثانى المسحوب غير جيد (معاب) | = D
نلاحظ أن :

$$P(D/A_1) = \frac{1}{4}$$

$$P(D/A_2) = 0$$

ولذلك فإن الاحتمال المطلوب يحسب بالعلاقة :

$$P(D) = P(D/A_1) . P(A_1) + P(D/A_1) . P(A_2)$$

وهذا الاحتمال يساوى :

$$= \frac{3}{7} . \frac{1}{4} = \frac{3}{28} = 0.107142857$$

تمرين (١٨)

ينتج مصنع للأنابيب الإلكترونية نوعا من اللمبات ، وقد لوحظ أن متوسط اللمبات المعيبة فى إنتاج آلة معينة هو 20% ما هو احتمال :

١) وجود لمبتين معابتين عند اختيار عشرة بصورة عشوائية ؟
٢) وجود أكثر من لمبتين معابتين عند اختيار عشرة بصورة عشوائية ؟
٣) وجود أكثر من خمس لمبات معابة عند اختيار عشرة بصورة عشوائية ؟

الحل

لنرمز للحوادث :

A = | عدد اللمبات المعابة اثنتين |
B | عدد اللمبات المعابة أكثر من لمبتين |
C = | عدد اللمبات المعابة أكثر من خمس |

نلاحظ أن :

$$1)\ P(A) = \binom{10}{2}(0.2)^2 . (0.8)^8 = 0.0302$$

2) p(B) = 1 - (أن يوجد اثنتين أو أقل معابتين)

$$= 1 - \binom{10}{0}(0.2)^0 . (0.8)^{10} - \binom{10}{1}(0.2)^1 . (0.8)^9$$

$$- \binom{10}{2} . (0.2)^2 (0.8)^8$$

$$= 1 - 0.1074 - 0.2681 - 0.0302 = 0.594$$

وأخيرا فإن :

3) P(C) = P [6 عدد اللمبات المعابة] + P[7 عدد اللمبات المعابة] +
P [8 عدد اللمبات المعابة] + P[9 عدد اللمبات المعابة] +
P [10 عدد اللمبات المعابة]

$$= \binom{10}{6}(0.2)^6 (0.8)^4 + \binom{10}{7}(0.2)^7 (0.8)^3 +$$

$$= \binom{10}{8}(0.2)^8 (0.8)^2 + \binom{10}{9}(0.2)^9 (0.8)^1 +$$

$$\binom{10}{10}(0.2)^{10} (0.8)^0$$

$$= 0.00637$$

تمرين (١٩)

تتألف مجموعة عناصر من مائة عنصر تخضع للفحص الختبري الجزئي ، ونشترط على عدم صلاحية هذه المجموعة إذا توافر فيها عنصر غير سليم على الأقل بين كل خمسة عناصر مراقبة . ما هو احتمال أن تكون مجموعة العناصر السابقة غير صالحة إذا احتوت على 5% من العناصر غير السليمة ؟

الحل

بفرض A = المجموعة صالحة عندئذ يكون المطلوب حساب P(A) ومن المعلوم

أن :
$$P(\acute{A}) = 1 - P(A) = 1 - q$$

لنسحب q ، إذا رمزنا للحوادث { العنصر المراقب ذى الرقم K سليم } = A_k .
عندئذ نجد أن :

$$A = A_1 \cap A_2 \cap \dots A_5$$

والملاحظ أن :

$$P(A_1) = \frac{95}{100}$$

لأن المائة عنصر تحوى 5 عناصر غير سليمة . وبعد تحقق الحادث A_1 يبقى 99 عنصرا من بينها أربع وتسعون عنصرا سليما ولذلك فإن :

$$P(A_2/A_1) = \frac{94}{99}$$

وبنفس الطريقة نجد أن : $P(A_3/A_1 \cap A_2) = \frac{93}{98}$, $P(A_4/A_1 \cap A_2 \cap A_3) = \frac{92}{97}$.

$$P(A_5/A_1 \cap A_2 \cap A_3 \cap A_4) = \frac{91}{96}$$

وباستخدام النظرية (١,١٢) فإننا نجد أن :

$$q = P(A) = P(A_1 \cap A_2 \cdots \cap A_5) = P(A_1) \cdot P(A_2/A_1) \cdot$$
$$P(A_3/A_1 \cap A_2) \cdot P(A_4/A_1 \cap A_2 \cap A_3) \cdot P(A_5/A_1 \cap A_2 \cap A_3 \cap A_4)$$

$$q = 0.77$$

$$P(A) = 0.23$$

تمارين عامة

(١) عين عناصر فضاءات العينة التالية :

أ ــ مجموعة الأعداد الصحيحة المحصورة بين الواحد والخمسين والتى تقبل القسمة على سبعة .

ب ــ $S = \{x | x^2 + x - 6 = 0\}$

جـ ــ مجموعة النتائج الممكنة عند إلقاء حجر نرد وقطعة نقود فى آن واحد

د ــ $S = \{x | 2x - 4 = 0, x > 5\}$

(٢) عين الحوادث المتساوية فيما يلى :

$A = \{1,3\}$

B = { x | x يمثل الوجه الذى ظهر عند إلقاء حجر نرد }

$C = \{ x | x^2 - 4x + 3 = 0 \}$

D = { x | x هو عدد مرات الصورة التى ظهرت عند إلقاء ست قطع نقود }

(٣) ألقينا حجرى نرد متماثلين ومتوازيين أحدهما أخضر والآخر أحمر ، والمطلوب :

أ ــ تحديد عناصر فضاء العينة S .

ب ــ تحديد عناصر الحادث { مجموع الوجهين اللذين ظهرا أقل من خمسة } = A

جـ ــ تحديد عناصر الحادث { ظهر على أحد الوجهين العدد ستة } = B

د ــ تحديد عناصر الحادث { ظهر على الحجر الأخضر الوجه اثنان } = C

هـ ــ ارسم مخطط فين لتوضيح العلاقة بين الحوادث S.A.B.C

(٤) تتألف تجربة من إلقاء قطعة نقود ثم إعادة هذا الإلقاء إذا ظهر وجه صورة . فإذا علمت أنه قد ظهر كتابة على القطعة الأولى الملقاة ، وأننا ألقينا بعد ذلك حجر نرد فالمطلوب :

أ ــ تحديد عناصر فضاء العينة S .

ب ــ تحديد نقاط العينة الموافقة للحادث A = { ظهور عدد أقل من 4 على حجر النرد } = A

جـ ــ تحديد نقاط العينة الموافقة للحادث { ظهور الكتابة مرتين } = B

(٥) اختير أربعة طلاب من جامعة الملك عبد العزيز بجدة من كليتى الهندسة والطب ،

ثم صنفوا بحسب الفرع الذى ينتمون إليه . أوجد عناصر فضاء العينة S باستخدام الرمز E لطالب الهندسة ، والرمز M لطالب الطب ، بفرض S_2 يمثل عدد طلاب كلية الهندسة المختارين ، فالمطلوب تحديد عناصر فضاء العينة S_2 .

(٦) بفرض أن فضاء العينة :

S = { صوديوم ، بوتاسيوم ، نيتروجين ، يورانيوم ، أكسجين ، زنك ، فضة }

وأن الحوادث :

A = { فضة ، صوديوم ، زنك }

B = { صوديوم ، نيتروجين ، بوتاسيوم }

C = { أكسجين ، يورانيوم ، زنك }

فالمطلوب تحديد عناصر الحوادث التالية :

أ : $\acute{A}$

ب : $A \cup C$

جـ : $(A \quad \acute{B}) \cup \acute{C}$

د : $\acute{B} \cap \acute{C}$

هـ : $A \cap B \cap C$

(٧) إذا علمت أن الحادثين :

$$A = \{ x \mid 1 < x < 9 \}, B = \{ y \mid y < s \}$$

فأوجد الحادثين :

$$A \cup B. \ A \cap B$$

(٨) بكم طريقة يمكن أن يصطف خمسة أشخاص لدى صعودهم إلى الباص ؟

(٩) ما هو عدد التباديل المختلفة التى يمكن تشكيلها من كلمة « الإحصاء » ؟

(١٠) بكم طريقة يمكن أن يجلس أربعة أطباء وثلاثة مهندسين فى صف إذا كان عليهم أن يجلسوا بالتناوب ؟

(١١) كم لجنة مؤلفة من ثلاثة أشخاص يمكن تشكيلها من ضمن خمسة مهندسين وثلاثة أطباء فى الحالات التالية :

أ ــ بدون تحديد نوع اللجان .

ب ـــ إذا ضُمِّنت كل لجنة مهندسين وطبيباً .
جـ ـــ إذا ضُمِّنت كل لجنة طبيبا ومهندسا ، وكان هناك طبيب معين موجود فى كل لجنة .

(١٢) يحتوى صندوق على ثلاث تفاحات حمر وأربع خضر وخمس صُفر . بكم طريقة يمكن سحب ست تفاحات من الصندوق إذا ضمنت تفاحتين من كل لون ؟

(١٣) سحبنا ورقتين من ورق اللعب المؤلف من 52 ورقة وذلك بصورة عشوائية ، ما هو احتمال أن تكون كلتا الورقتين أكبر من 2 وأقل من 9 ؟

(١٤) إذا علمت أن احتمال الحادثين B, A المستقلين هما على الترتيب $P(A) = 0.4, \quad P(B) = 0.5$ فما هو :

$P(A \cup B)$ و $P(A')$ و $P(A' \cup B)$

(١٥) يوجد سيارتان مستقلتان لإطفاء الحرائق فى مركز إطفاء البغدادية فى مدينة جدة ، فإذا علمت أن احتمال أن تكون سيارة معينة منهما جاهزة لإطفاء حريق هو 0.99 فالمطلوب :

أ ـــ حساب احتمال ألا تكون أية سيارة منهما جاهزة لإطفاء الحريق عندما نحتاج لها .

ب ـــ حساب احتمال أن تكون سيارة حريق جاهزة عند احتياجنا لها .

(١٦) إذا علمت أن احتمال أن يبقى أحمد على قيد الحياة خلال العشرين عاما المقبلة هو 0.6 ، وأن احتمال أن تبقى زوجته على قيد الحياة خلال العشرين عاما المقبلة هو 0.9 ، فما هو احتمال ألا يبقى أحمد أو زوجته على قيد الحياة خلال نفس الفترة ؟

(١٧) إذا علمت أن احتمال أن يفتح عمار التليفزيون على برنامج معين للأطفال هو 0.4 ، واحتمال أن تفتح شقيقته التليفزيون على نفس البرنامج هو 0.5 ، وأن احتمال أن يفتح عمار التليفزيون على البرنامج إذا علم أن شقيقته قد فتحت على نفس البرنامج هو 0.7 فالمطلوب حساب :

أ ـــ احتمال أن يفتح كلاهما على نفس البرنامج .
ب ـــ احتمال أن تفتح البنت على البرنامج علما أن عمار قد فتح عليه .
جـ ـــ احتمال أن يفتح أحدهما على الأقل على نفس البرنامج .

(١٨) يحوى كيس أربع كرات بيض وثلاث سود ، ويحوى كيس آخر ثلاث كرات بيض وخمس سود . سحبنا بصورة عشوائية كرة من الكيس الأول ووضعناها فى الثانى . ما هو احتمال سحب كرة سوداء من الكيس الثانى ؟

(١٩) ثلاثة صناديق تحتوى على كرات ملونة موزعة على النحو التالى :

اللون	الصندوق الأول	الثانى	الثالث
أحمر	12	4	3
أبيض	9	11	4
أزرق	5	6	8

سحبنا صندوقا بصورة عشوائية ، ثم سحبنا منه كرة بصورة عشوائية أيضا ، فلاحظنا أنها حمراء . ماهو احتمال أن يكون الصندوق المسحوب هو الثانى ؟

الفصل الثاني

المتغيرات العشوائية

■ المتغير العشوائي ■ التوزيع الاحتمالي المتقطع ■ توزيع الاحتمال المستمر ■ التوزيعات التجريبية ■ توزيع الاحتمال المشترك ■ التوقع الرياضي ■ قوانين التوقع الرياضي ■ التوقعات الرياضية الخاصة (التباين–المتغير) ■ خواص التباين ■ نظرية تشبيشيف ■ تمارين محلولة ■ تمارين عامة .

(٢,١) المتغير العشوائى
The random variable

من الواضح أن التجربة الإحصائية هى عملية تؤدى إلى قياس أو ملاحظات ، وأن معظم التجارب ذات الأهمية التجريبية تقودنا إلى قياسات عديدة تتغير من نقطة عينة إلى أخرى ، وبالتالى فإن هذا القياس يدعى بالمتغير العشوائى .

تعريف (٢,١) المتغير العشوائى :

المتغير العشوائى هو دالة عددية تأخذ قيما حقيقية تتحدد بوساطة كل عنصر من عناصر فضاء العينة S .

سنرمز للمتغير العشوائى بأحد الأحرف الكبيرة .. X,Y,Z, ، أما القيم التى يأخذها هذا المتغير فى نقطة عينة معينة فسنرمز لها بأحد الحروف الصغيرة ... x,y,z, . بفرض أن التجربة الإحصائية تمثل إحصاء عدد الصور التى نحصل عليها لدى إلقاء قطعة نقود مرتين . فمن الواضح عندئذ أن عدد الصور هذا ما هو إلا متغير عشوائى نرمز له بالرمز X ، وهذا المتغير يأخذ القيم 0,1,2 عند نقاط العينة الأربعة فى فضاء العينة :

$$S = \{ HH, HT, TH, TT \}$$

مثال (٢,١)

يحتوى وعاء على ثلاث كرات حمر وخمس بيض . ولنسحب كرتين من هذا الوعاء بصورة عشوائية على التتالى ، وبدون إعادة الكرة المسحوبة إلى الوعاء . من الواضح أنه إذا رمزنا بـ X لعدد الكرات الحمر المسحوبة فى هذه التجربة وبالرمز x لقيم هذا المتغير العشوائى فإننا نجد أن :

الجدول (٢,١)

X	الحوادث البسيطة
3	RRR
2	RWR
2	WRR
2	RRW
1	RWW
1	WRW
1	WWR
0	WWW

مثال (٢,٢)

فى تجربة إلقاء حجرى نرد دفعة واحدة ، إذا رمزنا بـ Y لأكبر العددين اللذين ظهرا على وجهى هذين الحجرين ، أى أنه من أجل أى نقطة $(x,y) \in S$ فإن $Y(x,y) = \max(x,y)$. عندئذ تكون قيم المتغير Y

$$Y: 1,2,3,4,5,6$$

مثال (٢,٣)

لنرمز لعدد القذفات اللازمة للحصول على وجه الصورة لأول مرة عند إلقاء قطعة نقود بالرمز Y من الواضح أن قيم Y هى .

$$Y: 1,2,3,4,5...$$

إذا أمعنا النظر فى المثالين (٢,١) ، (٢,٢) فإننا نجد أن فضاء العينة فى كل منهما يتألف من عدد منته من النقاط ، أما فضاء العينة فى المثال (٢,٣) فيتألف من عدد غير منته من نقاط العينة إلا أنه معدود .

تعريف (٢,٢) متغير عشوائى منقطع Discrete random variable

إذا حوى فضاء عينة ما على عدد منتهٍ أو غير منتهٍ ولكنه معدود من نقاط العينة ، قلنا إن هذا الفضاء منقطع ، ونسمى المتغير العشوائى المعرف على هذا الفضاء بمتغير عشوائى منقطع .

لنسأل الآن أنفسنا عن فضاء العينة الموافق لتجربة قياس المعدل اليومى لهطول المطر فى مدينة جدة خلال شهر معين ، إذا استخدمنا (مسطرة مدرجة) لهذا القياس ، فإن معدل الهطول سيوافق نقطة على هذه المسطرة والتى يمكن النظر إليها على أنها نقطة من محور موجه . وفضاء العينة يمكن إذاً أن يكون أى نقطة من مجال على هذا المحور . ومن المعلوم أنه يوجد من النقاط فى أى مجال من محور موجه مهما كان صغيراً عدداً غير معدود من النقاط .

تعريف (٢,٣) متغير عشوائي مستمر Continuous random variable

إذا حوى فضاء العينة فى تجربة معينة على عدد غير معدود وغير محدود من النقاط قلنا عنه أنه فراغ عينة مستمر . كما نسمى المتغير العشوائى المعرف على هذا النوع من الفراغات بمتغير عشوائى مستمر .

نلاحظ أن الأوزان والأطوال والقوى كلها أمثلة على المتغيرات العشوائية المستمرة ، وقياسات مثل هذه المتغيرات تشكل بدون انقطاع نقاط على محور موجه أو مجالات بين قياس وآخر .

(٢,٢) التوزيع الاحتمالى المنقطع

Diserete Probability Distribution

إن العدد الممثل للوجه الذى يظهر عند إلقاء حجر نرد والذى نرمز له بالرمز X هو متغير عشوائى يأخذ القيم 1,2,3,4,5,6 باحتمالات متساوية . كل منها يساوى العدد $\frac{1}{6}$. كما نلاحظ أن المتغير العشوائى الممثل لعدد الصور التى نحصل عليها لدى إلقاء قطعة نقود أربع مرات متتالية سيأخذ القيمة 3 باحتمال قدرة $\frac{1}{4}$. ونلاحظ أيضا فى المثال (٢,٢) أن المتغير العشوائى Y يفترض القيم 1,2,3,4,5,6 باحتمالات موافقة قدرها $\frac{1}{36}$ ، $\frac{3}{36}$ ، $\frac{5}{36}$ ، $\frac{7}{36}$ ، $\frac{9}{36}$ ، $\frac{11}{36}$ على الترتيب :

ومن المناسب فى كثير من الحالات أن نمثل الاحتمال الموافق لقيمة معينة من قيم المتغير العشوائي المفروض على شكل دالة بهذه القيمة ، نرمز لها بالرمز f(x) وعلى هذا

نكتب $f(x) = P[X = x]$. فمثلا نجد فى المثال (٢,٢) أن :

$$f(1) = P[X = 1] = P[(1,1)] = \frac{1}{36}$$

$$f(2) = P[X = 2] = P[(2,1), (2,2), (1,2)] = \frac{3}{36}$$

وهكذا ....

تعريف (٢,٤) التوزيع الاحتمالى Probability distribution

تسمى الدالة $f(x)$ بتوزيع الاحتمال للمتغير العشوائى المنقطع X إذا حققت الشروط الثلاثة التالية :

1) $f(x) \geq 0$

2) $\sum_{x} f(x) = 1$

$P(x) = P[X = x]$

وذلك لجميع قيم x .

مثال (٢,٤)

أوجد توزيع الاحتمال للمتغير العشوائى Y الوارد فى المثال (٢,٢) ؟

الحل

من الواضح (انظر المثال ١,٩) أن عدد نقاط العينة عند إلقاء حجرى النرد هو 36 نقطة ، وكما نعلم فإن قيم المتغير Y هى { 1,2,3,4,5,6 } ولذلك :

$$f(1) = P(Y = 1) = P[(1,1)] = \frac{1}{36}$$

$$f(2) = P(Y = 2) = P[\{(1,2), (2.2), (2,1)\}] = \frac{3}{36}$$

$$f(3) = P(Y = 3) = P[\{(1,3), (3,1), (3,3), (2,3), (3,2)\}] = \frac{5}{36}$$

$$f(4) = P(Y = 4) = P[\{(1,4), (4,1) (2,4), (4,2), (3,4), (4,3), (4,4)\}] = \frac{7}{36}$$

وأخيراً فإن :

$$f(5) = P(Y = 5) = \frac{9}{36}, \; f(6) = P(Y = 6) = \frac{11}{36}$$

يمكن تنظيم قيم المتغير العشوائى والاحتمالات الموافقة بجدول كالتالى :

Y	1	2	3	4	5	6	
f(y)	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{5}{36}$	$\frac{7}{36}$	$\frac{9}{36}$	$\frac{11}{36}$	$\sum_{Y_i} f(y_i) = 1$

مثال (٢,٥)

لنبحث عن توزيع الاحتمال لعدد مرات الصورة التى نحصل عليها لدى إلقاء قطعة نقود متماثلة ومتوازنة ثلاث مرات .

الحل

بالرجوع إلى المثال (١,١٨) نجد أن عدد نقاط فضاء العينة S هو 8 وأن هذه النقاط متساوية فى إمكانية وقوعها ، أما عدد أوجه الصورة التى نحصل عليها فيساوى : X: 0,1,2,3 ، كما أن الاحتمالات الموافقة لهذه القيم فهى :

$$f(0) = P[X = 0] = \frac{\binom{3}{0}}{8} = \frac{1}{8}$$

$$f(1) = P[X = 1] = \frac{\binom{3}{1}}{8} = \frac{3}{8}$$

$$f(2) = P[X = 2] = \frac{\binom{3}{2}}{8} = \frac{3}{8}$$

$$f(3) = P[X = 3] = \frac{\binom{3}{3}}{8} = \frac{1}{8}$$

والملاحظ فى هذا المثال أننا استخدمنا عدد الطرق التى يمكن بها تجزئة n شيئا إلى مجموعتين : تحتوى إحداهما على x شيئا ، والثانية على n - x شيئا آخر ، وهذا العدد ما هو إلا التوفيقة $\binom{n}{x}$. وهكذا نجد أن جدول توزع المتغير العشوائي X :

X	0	1	2	3	
f(x)	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$	$\sum_{X_i} f(x_i) = 1$

تعريف (٢,٥) التوزيع التراكمى Cumulative distribution

نعرف التوزيع التراكمى F(x) لمتغير عشوائى منقطع X له توزيع احتمال f(x) بالعلاقة :

$$F(x) = P[X \leq x] = \sum_{y \leq x} f(y)$$

مثال (٢,٦)

لنبحث عن التوزيع التراكمى للمتغير العشوائى X فى المثال (٢,٥)

الحل

نلاحظ أن :

$$F(0) = f(0) = \frac{1}{8}$$

$$F(1) = f(0) + f(1) = \frac{1}{8} + \frac{3}{8} = \frac{1}{2}$$

$$F(2) = f(0) + f(1) + f(2) = \frac{7}{8}$$

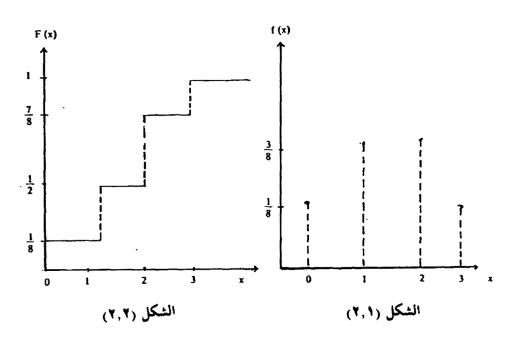
وهكذا نجد أن :

$$F(3) = f(0) + f(1) + f(2) + f(3) = 1$$

$$F(x) = \begin{cases} 0 & \text{for} \quad x < 0 \\ \frac{1}{8} & \text{for} \quad 0 \leq x < 1 \\ \frac{1}{2} & \text{for} \quad 1 \leq x < 2 \\ \frac{7}{8} & \text{for} \quad 2 \leq x \\ 1 & \text{for} \quad 3 \leq x \end{cases}$$

تمثل بيانياً كلا من دالة توزيع الاحتمال ، ودالة التوزيع التراكمى .

يبين الشكلين (٢،١) ، (٢،٢) على الترتيب الشكل البيانى لدالة توزيع الاحتمال ، ولدالة التوزيع التراكمى .



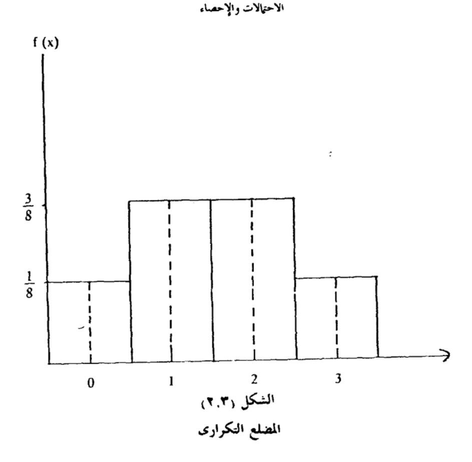
الشكل (٢،٢)

الشكل (٢،١)

ملاحظة هامة

فى كثير من الأحيان نقوم بدلاً من رسم النقاط $(x, f(x))$ بإنشاء مستطيلات كما فى الشكل (٢،٣) الموضح أدناه . هذه المستطيلات لها قواعد متساوية ومتمركزة فى النقاط الممثلة لقيم المتغير العشوائى ، أما ارتفاعاتها فتساوى قيم $f(x)$ الموافقة ويسمى الشكل (٢،٣) بالمضلع التكرارى لـ X ويوضح لنا المضلع التكرارى لـ X الوارد فى المثال (٢،٥) والموضح على الشكل (٢،٣) أن $P[X = x]$ ما هو إلا مساحة هذا المستطيل المتمركز فى النقطة X ذلك لأن عرض المستطيل هو الواحد حتى ولو لم تكن المستطيلات ذات قواعد عرضها الواحد فإننا نقوم بتعديل ارتفاع هذه المستطيلات لتبقى مساحات المستطيلات فى المضلع التكرارى مساوية للاحتمالات الموافقة لقيم المتغير X .

هذه الفكرة حول استخدام المساحات لتمثيل الاحتمالات ضرورية وهامة فى دراسة دوال التوزيع للمتغيرات العشوائية المستمرة .



الشكل (٢،٣)

المضلع التكرارى

(٢،٣) توزيع الاحتمال المستمر

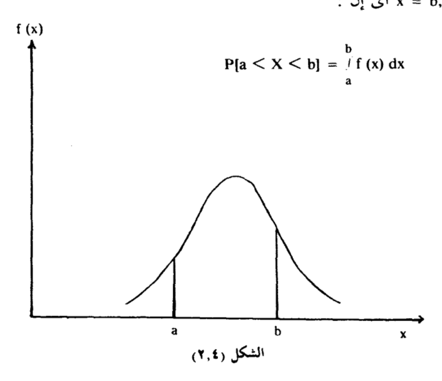
Continuous propability distribution

كما أشرنا فى التعريف (٢،٣) إلى أن قياسات المتغيرات المستمرة تشكل بدون انقطاع على محور موجه أو مجالات بين قياس وآخر ، وبما أنه لا يمكننا تخصيص احتمال لكل نقطة عينة ، فلابد من التفكير فى نموذج احتمالى مختلف عما رأيناه فى حالة المتغيرات المنقطعة . فلو عدنا إلى المثال (٢،٥) ، والشكل (٢،٣) الممثل للمضلع التكرارى فى هذا المثال ، وقمنا بزيادة عدد مرات إلقاء قطعة النقود بشكل كبير ، فعندئذ سيضيق عرض المجالات الجزئية ، وسيتغير المظهر العام للمضلع التكرارى فى اتجاه التخلص من مظاهر عدم الانتظام . وعندما يصبح عدد القياسات كبيراً جداً ، وعرض المجالات الجزئية صغيراً جداً فسيظهر التكرار النسبى وكأنه عمليا منحنى لنعدل المساحة الكلية تحت المضلع التكرارى هذا لتصبح مساوية الواحد .

لنفرض الآن المتحول العشوائى المستمر X يأخذ قيمة فوق محور الأعداد الحقيقية ضمن المجال (a, b) ، ولنوزع احتمالا قدره الواحد على طول هذا المجال إلى حد كبير ، كما

يوزع شخص حفنة من الرمال ، بحيث توافق كل ذرة رمل قياسا من قياسات المجتمع . وستتجمع ذرات الرمل أو القياسات ، ونقول عندئذ أن كثافة الاحتمال فى مثل هذه الأماكن أكبر منها فى أماكن أخرى ، وسيكون لهذه الكثافة قيمة غير سالبة (موجبة أو صفر) فى كل نقطة من نقاط المجال . فلو تصورنا أن هذه الكثافة تتغير من نقطة إلى أخرى وفق علاقة f(x) فإن منحنى الدالة f(x) هو منحنى الكثافة (density function) . ويظهر الشكل (٢,٤) توضيحا لدالة الكثافة ، وبما أن المساحة أسفل المنحنى f(x) يجب أن تساوى الواحد ، لذا فإن المساحة أسفل المنحنى وفوق المجال (a,b) تساوى الواحد .

وبشكل عام إذا كان X متغيراً عشوائياً مستمراً ، دالة كثافته الاحتمالية f(x) ، وإذا كان (a,b) مجالا ما على المحور الحقيقى ، عندئذ يكون احتمال أن يأخذ x قيمة ضمن هذا المجال P[a<X<b] مساوياً إلى المساحة تحت منحنى الكثافة f(x) بين الخطين الشاقوليين $x = b, x = a$ أى إن :



الشكل (٢,٤)

تعريف (٢,٦) دالة الكثافة الاحتمالية Probability density function

ليكن X متغيرا عشوائياً مستمراً معرفاً فوق المحور الحقيقى R . نسمى الدالة f(x) بدالة الكثافة الاحتمالية لهذا المتغير إذا حققت الشروط الثلاثة التالية :

$$1 - \quad f(x) \geq 0 \quad \text{for all} \quad x \in R$$

$$2 - \quad \int_{-\infty}^{+\infty} f(x)\, dx = 1$$

$$3 - \quad P[a < X < b] = \int_a^b f(x)\, dx$$

مثال (٢,٧)

ليكن X متغيراً عشوائياً مستمراً معرفاً على المجال (1,3) بالكثافة الاحتمالية :

$$f(x) = \frac{1}{2} \quad : \quad 1 < x < 3$$
$$= 0 \quad : \quad \text{فيما عدا ذلك}$$

لنرى إن كان f(x) يمثل دالة كثافة احتمالية أم لا ، ثم لنبحث عن $P(2<x<2.5)$. نلاحظ أولا أنه من أجل جميع قيم x فوق المجال المذكور فإن $f(x) = \frac{1}{2} > 0$

ثم إن

$$\int_{-\infty}^{+\infty} f(x)\, dx = \int_2^3 \frac{1}{2}\, dx = 1$$

وأخيراً فإن :

$$P(2 < X < 2.5) = \int_2^{2.5} \frac{1}{2}\, dx = \frac{1}{4}$$

مثال (٢,٨)

بفرض أن الدالة :

$$f(x) = \frac{2(1+x)}{27} : \quad 2 < x < 5$$
$$= \quad 0 \quad : \quad \text{فيما عدا ذلك}$$

تمثل دالة كثافة لمتغير عشوائي مستمر . احسب $P[3 < X < 4]$.

نلاحظ أن :

$$P[3 < x < 4] = \int_3^4 f(x)dx = \frac{2}{27}\int_3^4 (1 + x)\,dx = \frac{1}{3}$$

تعريف (٢,٧) التوزيع التراكمى The cumulative distribution

نقول بأن الدالة F(x) تمثل دالة التوزيع التراكمى لمتغير عشوائى مستمر X بالكثافة f(x) إذا حققت الدالة F(x) العلاقة الآتية :

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(u)\,du$$

من التعريف السابق نستنتج النتائج الهامة التالية :

1. $P(a < X < b) = F(b) - F(a)$

2. $\frac{dF(x)}{dx} = f(x)$

مثال (٢,٩)

لنفتش فى المثال (٢,٧) عن دالة التوزيع التراكمى F ، ثم لنحسب $P[2 < X h 3]$

الحل

من التعريف (٢,٧) نلاحظ أن :

(١) إذا كان $X \leq 1$ فإن $F(x) = 0$ لأنه لا توجد كثافة على المجال $(-\infty, 1)$

(٢) أما إذا كان $1 < X \leq 3$ فإن :

$$F(x) = \int_{-\infty}^{x} f(u)\,du = \int_{-\infty}^{1} f(u)\,du + \int_{1}^{x} f(u)\,du$$

$$= \int_{-\infty}^{1} 0,\,du + \int_{1}^{x} \frac{1}{2}\,du = \frac{x-1}{2}$$

(٣) وإذا كان $x > 3$ فإننا نجد أن :

$$F(x) = \int_{-\infty}^{1} f(u)du + \int_{1}^{3} f(u)du + \int_{3}^{x} f(u)du$$

$$= \int_{-\infty}^{1} 0.du + \int_{1}^{3} \frac{1}{2} du + \int_{3}^{x} 0.du = 1$$

وأخيرا فإننا نلاحظ أن :

$$F(x) = \begin{cases} 0 & \text{if } x \leq 1 \\ \frac{x-1}{2} & \text{if } 3 \geq x > 1 \\ 1 & \text{if } x > 3 \end{cases}$$

كذلك نلاحظ أن :

$$P(2 < x \leq 3) = F(3) - F(2) = \frac{3-1}{1} - \frac{2-1}{2} = 1 - \frac{1}{2} = \frac{1}{2}$$

(٢,٤) التوزيعات التجريبية Empirical distribution

يمكن وصف مجموعة من القياسات بإحدى طريقتين : تدعى الأولى بالطريقة البيانية ، أما الثانية فتسمى بالطريقة العددية وهى بديهية ، حيث تقوم بوصف القياسات دون تحليلها . سنشرح فيما يلى الطريقة البيانية بشىء من التفصيل :

الطريقة البيانية :

يمثل الجدول (٢,٢) الموضح أدناه معدلات ثلاثين طالبا من طلاب السنة الجامعية الأولى فى إحدى جامعات المملكة . بالقاء نظرة سريعة على هذا الجدول يتبين لنا أن أقل معدل هو 46.7 ، وأعلى معدل هو 96.7 والسؤال المطروح هل : القياسات الثمانى والعشرين الباقية تتوزع بصورة عادلة فى المجال (46.7,96.7) ، أم أنها تقع بالقرب من أطراف المجال السابق ؟ ولكى نجيب على هذا السؤال نقوم بتقسيم المجال السابق إلى عدد من المجالات الجزئية المتساوية يتوقف عددها على عدد القياسات الموجودة (عادة نختار من 5 إلى 20 مجالاً جزئياً يتناسب عددها مع عدد القياسات التى ندرسها) . فمثلا : يمكن أن نستخدم المجالات الجزئية (52.90,59.15) (59.15,65.40) (65.4,71.65) (46.65,52.90) إلخ ...

الجدول (٢,٢)

الجدول (٢,٢) يبين معدلات 30 طالبا فى إحدى جامعات المملكة

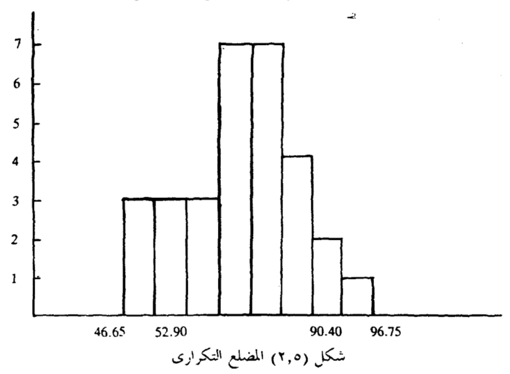
50.0	86.7	46.7	66.7	46.7
60.0	70.0	86.7	66.7	53.3
80.0	83.3	73.3	66.7	63.3
73.0	66.7	63.3	83.3	96.7
70.0	76.7	66.7	73.3	80.0
73.3	76.7	56.7	73.3	53.3

والمتتبع لهذه المجالات يلاحظ أن أى قياس لا يقع فوق نقاط التقسيم . لنصنف الآن القياسات الموجودة فى الجدول (٢,٢) كلا فى المجال الجزئى الذى يحويه ، يبين الجدول (٢,٣) هذا التصنيف :

الجدول (٢,٣)

رقم المجال الجزئى	حدود كل مجال جزئى	عدد القياسات التى تقع فى المجال الجزئى المقابل	عدد القياسات التى تقع فى المجال الجزئى المقابل	نسبة القياسات الواقعة ضمن كل مجال جزئى إلى العدد الكلى للقياسات التى تصنفها
1	64.56 - 52.90	III	3	3/30
2	52.90 - 59.15	III	3	3/30
3	59.15 - 65.40	III	3	3/30
4	65.40 - 71.65	VII	7	7/30
5	71.65 - 77.90	VII	7	7/30
6	77.90 - 84.15	IV	4	4/30
7	48.15 - 90.40	II	2	2/30
8	90.40 - 96.75	I	1	1/30
		المجموع	n = 30	1

فنسمي النسبة المذكورة فى العمود الخامس بالتكرار النسبى (relative frequency) للمجال الجزئى الموافق . هذا ويمكن تمثيل الجدول الناتج بيانيا على شكل مضلع تكرارى حيث نقوم بإنشاء مستطيل (فوق كل مجال جزئى) ارتفاعه متناسب مع عدد القياسات الواقعة ضمن هذا المجال . يبين الشكل (٢,٥) مثل هذا التمثيل .



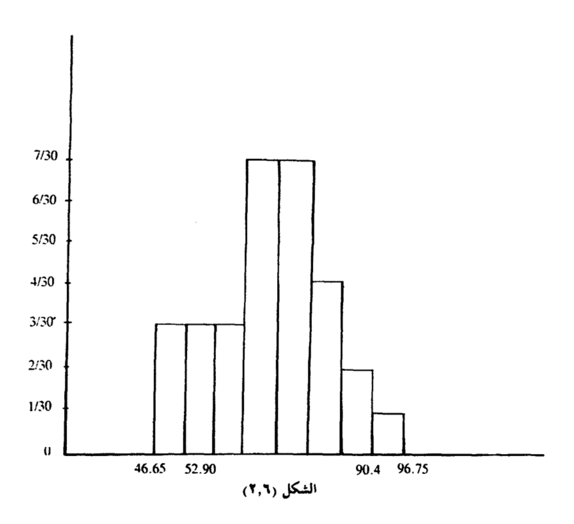
شكل (٢,٥) المضلع التكرارى

غالبا ما يكون من الأنسب تعديل المضلع التكرارى بحيث نختار وحدة التقسيمات على المحور الشاقولى . فإذا اعتبرنا أن ارتفاع كل مستطيل ننشئه فوق كل مجال جزئى مساوياً للتكرار النسبى الموافق لهذا المجال كما فى الشكل (٢,٦) فإنه من النادر عندئذ التمييز بين المضلعين .

بالعودة إلى مضلع التكرار نلاحظ أن 14 طالبا قد حصلوا على معدلات أكبر من 71.65 . كما نلاحظ أيضا أن هذه النسبة تساوى النسبة المئوية للمساحة الواقعة إلى يمين 71.65 . وبما أن 14 طالبا قد حصلوا كل منهم على معدلا أكبر من 71.7 فإننا نقول أن هناك 14 فرصة من أصل 30 ، ونقول أيضا بأن احتمال الحصول على مثل هذا المعدل هو $\frac{14}{30}$.

ولو عدنا إلى المجتمع (مجموعة طلبة السنة الأولى) الذى سحبنا منه هذا العدد من الطلاب فما هى نسبة الطلاب الذين نالوا معدلا أكبر من 71.7 ؟ نلاحظ أن الجواب لهذا السؤال يكمن فى حساب نسبة المساحة الكلية (من مضلع التكرار النسبى الخاص بالمجتمع) الواقعة إلى يمين 71.65 . ونظرا لعدم وجود مثل هذا المضلع أمامنا فإننا سنضطر للقيام باستنتاج . ولابد قبل كل شىء من تقدير النسبة الحقيقية من المجتمع على أساس المعلومات التى تقدمها العينة (درجات الطلاب الثلاثين) . ومن المعقول أن يكون تقديرنا هنا هو $\frac{14}{30}$ أو 47% .

لنفرض أننا نود التعبير عن احتمال أن يكون لطالب انتقيناه من المجتمع بطريق الصدفة معدلا أكبر أو يساوى 71.7 .



الشكل (٢,٦)

يمكن أن نقول (وبدون معرفة مضلع التكرار النسبى للمجتمع) أن مضلع المجتمع مشابه لمضلع العينة ، وأن $\frac{14}{30}$ من قياسات المجتمع على وجه التقريب قد تكون أكبر أو تساوى 71.7 . ومن الطبيعى أن نرتكب فى عملية التقدير هذه خطأ . وسندرس مستقبلا كيف نضع حدوداً لمثل هذا الخطأ المرتكب فى التقدير .

يدعى مضلع التكرار النسبى بالتوزيع التكرارى ، لأنه يبين الطريقة التى تتوزع فيها المعلومات على طول محور الفواصل . والملاحظ أن المستطيلات المقامة فوق كل مجال جزئى خاضعة لتفسيرين . فمن جهة تمثل نسبة الملاحظات الواقعة فى مجال جزئى معين ، ثم إنه عند سحب قياس من القياسات وبشكل عشوائى يمكننا اعتبار مساحة المستطيل المقام فوق مجال جزئى معين ممثلة لاحتمال وقوع ذلك القياس ضمن هذا المجال الجزئى . والمهم فى مضلع التكرار للعينة هو أن هذا المضلع يمدنا بمعلومات حول المضلع التكرارى للمجتمع ، والشىء المتوقع أن يكون مضلع التكرار للمجتمع والعينة متشابهين ، وأن تزداد درجة التشابه هذه كلما ازداد حجم العينة ، وعندما تزداد العينة لتصبح متطابقة مع المجتمع المدروس فعندئذ يتطابق المضلعان التكراريان .

(٢,٥) توزيع الاحتمال المشترك Joint probability distribution

لقد اقتصرت دراستنا للمتغيرات العشوائية فى هذا الفصل على فضاء عينة واحدة . وفى هذا الفضاء فرضنا أن نتائج تجربة إحصائية هى قيم لمتغير عشوائى وحيد . ومع ذلك فإنه قد نكون مضطرين لاستخدام العديد من المتغيرات العشوائية فى وقت واحد . فمثلا لو طلب إلينا فى تجربة سحب ثلاث أوراق من ورق اللعب المكون من 52 ورقة حساب احتمال أن يكون بين الأوراق المسحوبة ورقة دينارى وورقتين كبة لكُنَّا مضطرين إلى اعتبار المتغير X ممثلا لعدد أوراق الدينارى المسحوبة و Y لعدد أوراق الكبة المسحوبة فى نفس التجربة .

وبشكل عام إذا كان Y,X متغيرين عشوائيين منقطعين ، وإذا رمزنا بـ $f(x,y)$ لاحتمال أن يأخذ المتغير X القيمة x والمتغير Y القيمة y أى إن :

$$P[X = x, Y = y] = f(x,y)$$

فعندئذ تكون هذه الدالة ممثلة لاحتمال أن يفترض المتغير X القيمة x ، والمتغير Y القيمة y . ونطلق على هذه الدالة اسم دالة توزيع الاحتمال المشترك . فمثلا فى المثال السابق الذى سقناه عند سحب ثلاث أوراق لعب يمثل $f(1,2)$ الاحتمال المطلوب .

تعريف (٢,٨) دالة توزيع الاحتمال المشترك Joint probability distribution function

نسمى الدالة $f(x,y)$ بدالة توزيع الاحتمال المشترك للمتغيرين العشوائيين المنقطعين X,Y إذا حققت الشروط الثلاثة التالية :

1. $f(x,y) \geq 0$ for all (x,y)

2. $\sum_{x}\sum_{y} f(x,y) = 1$

3. $P[(x, y) \in A] = \sum\sum_{A} f(x,y)$

وذلك من أجل أى منطقة A فى المستوى xy .

لايجاد جميع قيم دالة توزيع الاحتمال المشترك لمتغيرين عشوائيين منقطعين Y , X قيمهما $(y_1,\ldots, Y_n), (X_1, \ldots, x_n)$ على الترتيب نسوق الجدول التالى :

الجدول (٢,٤) جدول التوزيع المشترك لمتغيرين منقطعين

Y \ X	x_1	x_2	x_3	...	x_n	
y_1	$f(x_1,y_1)$	$f(x_2,y_1)$	$f(x_3,y_1)$	.....	$f(x_n,y_1)$	$\sum_{i=1}^{n} f(x_i,y_1)$
y_2	$f(x_1,y_2)$	$f(x_2,y_2)$	$f(x_3,y_2)$	.....	$f(x_n,y_2)$	$\sum_{i=1}^{n} f(x_i,y_2)$
y_3	$f(x_1,y_3)$	$f(x_2,y_3)$	$f(x_3,y_3)$	.....	$f(x_n,y_3)$	$\sum_{i=1}^{n} f(x_i,y_3)$
.....	.....	.....	.....	.....	.....	.....
y_n	$f(x_1,y_n)$	$f(x_2,y_n)$	$f(x_3,y_n)$	.....	$f(x_n,y_n)$	$\sum_{i=1}^{n} f(x_i,y_n)$
	$\sum_{i=1}^{n} f(x_1,y_i)$	$\sum_{i=1}^{n} f(x_2,y_i)$	$\sum_{i=1}^{n} f(x_3,y_i)$	...	$\sum_{i=1}^{n} f(x_n,y_i)$	

حيث يمثل العدد $f(x_i, y_j)$ احتمال أن يفترض المتغير X القيمة x_i ، والمتغير Y القيمة y_j أى أن :

$$f(x_i, y_j) = P[X = x_i, Y = y_j]$$

مثال (٢,١٠)

لنفرض فى تجربة إلقاء حجرى نرد متماثلين ومتوازيين أن X يمثل القيمة العظمى للعددين اللذين ظهرا ، وأن Y يمثل مجموع العددين . نلاحظ أن فضاء العينة مؤلف من 36 نقطة عينة . انظر المثال (١,١٠) . أما قيم المتغيرين Y,X فهى على الترتيب :

$X = \{ 1, 2, 3, 4, 5, 6 \}$

$Y = \{ 2, 3, 4, \dots , 12 \}$

أما توزيع الاحتمال المشترك لهذين المتغيرين المنقطعين فيوضحه الجدول التالى :

جدول (٢,٥)

Y \ X	1	2	3	4	5	6	
2	1/36	0	0	0	0	0	1/36
3	0	2/36	0	0	0	0	2/36
4	0	1/36	2/36	0	0	0	3/36
5	0	0	2/36	2/36	0	0	4/36
6	0	0	1/36	2/36	2/36	0	5/36
7	0	0	0	2/36	2/36	2/36	6/36
8	0	0	0	1/36	2/36	2/36	5/36
9	0	0	0	0	2/36	2/36	4/36
10	0	0	0	0	1/36	2/36	3/36
11	0	0	0	0	0	2/36	3/36
12	0	0	0	0	0	1/36	1/36
	1/36	3/36	5/36	7/36	9/36	11/36	

والحقيقة أن قيم دالة توزيع الاحتمال المشترك f(x,y) قد تم حسابها بواسطة العلاقة :

$$f(x,y) = P[X = x, Y = y]$$

وعلى سبيل المثال نلاحظ أن :

$f(2,4) = P[X = x, Y = y]$

$= P$ [مجموع الوجهين اللذين ظهرا 4 ، وأكبرهما 2]

$= P[(2,2)]$

$= \frac{1}{36}$

وهى نفس القيمة التى نلاحظها فى الجدول (٢,٥) فى العمود الثانى والسطر الرابع كذلك فإن :

$f(3,4) = P[X = 3, Y = 4]$

= [أكبر العددين اللذين ظهرا ثلاثة ، ومجموعهما أربعة]

$= P[(1,3), (3,1)] = \frac{2}{36}$

وهى القيمة التى نجدها فى العمود الثالث والسطر الرابع .

تعريف (٢,٩) دالة الكثافة المشتركة لمتغيرين عشوائيين مستمرين

نسمى الدالة f(x,y) بدالة الكثافة المشتركة لمتغيرين عشوائيين مستمرين Y,X إذا حققت الشروط الثلاثة التالية :

1) $f(x, y) \geq 0.$ for all (x,y)

2) $\int\int_{-\infty}^{+\infty} f(x, y)\, dx\, dy = 1$

3. $\int\int_{A} f(x,y)\, dx\, dy = P[(X, Y) \in A]$

وذلك مهما تكن المنطقة A من المستوى XY.

مثال (٢,١١)

لنفرض أن دالة الكثافة المشتركة للمتغيرين المستمرين (X,Y) هى من الشكل :

$$f(x,y) = \begin{cases} \frac{1}{8}(6-x-y) & : \quad 0 < x < 2 \\ & \quad\;\; 2 < y < 4 \\ 0 & : \quad \text{فيما عدا ذلك} \end{cases}$$

بما أن دالة التوزيع التراكمى F لهذه الثنائية العشوائية تتحدد بالعلاقة :

$$F(x,y) = \int_{u=-\infty}^{x} \int_{v=-\infty}^{y} f(u,v)\, du\, dv$$

فعندئذ نجد أنه إذا كان $x = 1$, $y = 3$ فعندئذ يكون :

$$F(1,3) = \int_{x=0}^{1} \int_{y=2}^{3} \frac{1}{8}(6-x-y)\, dx\, dy =$$

$$F(1,3) = \frac{1}{8}\int_{x=0}^{1} \left(6y - xy - \frac{y^2}{2} \right)_{y=2}^{y=3} dx$$

$$= \frac{1}{8}\int_{x=0}^{1} \left(\frac{7}{2} - x \right) dx = \frac{3}{8}$$

ملاحظة

إذا علمت دالة توزيع الاحتمال المشتركة $f(x,y)$ للمتغيرين X, Y فعندئذ يمكن معرفة توزيع الاحتمال لكل من Y, X كلا على حدة بوساطة العلاقتين التاليتين :

$$g(x) = \sum_{y} f(x,y)$$

$$h(y) = \sum_{x} f(x,y)$$

وذلك إذا كان المتغيران منقطعين .

أما إذا كان المتغيران مستمرين فإننا نستخدم العلاقتين التاليتين :

$$g(x) = \int_{-\infty}^{+\infty} f(x, y)\, dy$$

$$h(y) = \int_{-\infty}^{+\infty} f(x, y)\, dx$$

ونسمى كلا من الدالتين h(y) و g(x) بالتوزيع الهامشى (marginal distribution) لكل من المتغيرين X, Y على الترتيب .

وللتأكد من أن g(x), h(y) يمثل كلا منهما توزيع احتمال ، علينا أن نتأكد بسهولة من تحقق شروط التعريف (٢,٤) ، أو التعريف (٢,٦) على التتالى حسبما يكون المتغير منقطعا أو مستمراً على الترتيب . وبالحقيقة نلاحظ أن :

$$\int_{-\infty}^{+\infty} f(x)\, dx = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y)\, dy\, dx = 1.$$

كذلك فإن :

$$P[a < X < b] = P[a < X < b, -\infty < Y < +\infty]$$

$$= \int_{x=a}^{b} \int_{y=-\infty}^{+\infty} f(x, y)\, dy\, dx$$

$$= \int_{a}^{b} g(x)\, dx$$

ملاحظة هامة

بالعودة إلى الجدول (٢,٤) الممثل للتوزيع المشترك لمتغيرين منقطعين ، نلاحظ أن العمود الأخير فى هذا الجدول يمثل قيم التوزيع الهامشى للمتغير Y . فمثلا نلاحظ أن :

$y_1 \xrightarrow{\text{يقابلها}} \sum_{x_i} f(x_i, y_1)$

$y_2 \xrightarrow{\text{يقابلها}} \sum_{x_i} f(x_i, y_2)$

وهكذا ، كما نلاحظ أن عناصر السطر الأخير فى نفس الجدول تمثل قيم التوزيع الهامشى للمتغير X ، وعلى سبيل المثال نلاحظ أن القيم x_1, x_2, x_3 يقابلها على الترتيب القيم :

$$\sum_{j=1}^{n} f(x_1, y_j) \quad , \quad \sum_{j=1}^{n} f(x_2, y_j) \quad , \quad \sum_{j=1}^{n} f(x_3, y_j)$$

مثال (٢,١٢)

لنبحث عن قيم التوزيعات الهامشية لكل من Y, X فى المثال (٢,١٠) ، نلاحظ أن التوزيع الاحتمالى للمتغير X فى المثال (٢,١٠) يعطى بالجدول التالى :

X	1	2	3	4	5	6
P	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{5}{36}$	$\frac{7}{36}$	$\frac{9}{36}$	$\frac{11}{36}$

كذلك نلاحظ أن توزيع الاحتمال بالنسبة للمتغير Y يمثله الجدول :

Y	2	3	4	5	6	7	8	9	10	11	12
P	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

مثال (٢,١٣)

لنبحث عن دالة الكثافة الاحتمالية لكل من المتغيرين المستمرين Y, X فى المثال (٢,١١)

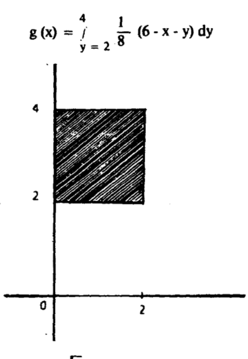
الحل

نلاحظ أن :

$$g(x) = \int_{-\infty}^{+\infty} f(x, y)\, dy$$

ومن عبارة الكثافة المشتركة فى المثال (٢,١١) نجد أن الكثافة الاحتمالية مجمعة فى القسم المظلل من الشكل المرفق ، أما خارج المربع المظلل ، فإن الكثافة معدومة وعلى هذا فإن :

$$g(x) = \int_{y=2}^{4} \frac{1}{8}(6 - x - y)\, dy$$



ومنه :

$$g(x) = \begin{cases} \frac{1}{8}(6 - 2x) & : \quad 90 < x < 2 \\ 0 & : \quad \text{فيما عدا ذلك} \end{cases}$$

كذلك فإن :

$$h(y) = \int_{-\infty}^{+\infty} f(x, y)\, dx = \int_{x=0}^{2} \frac{1}{8}(6 - x - y)\, dx$$

$$h(y) = \begin{cases} \frac{5 - y}{4} & : \quad 2 < y < 4 \\ 0 & : \text{فيما عدا ذلك} \end{cases}$$

لنفرض أن X, Y متغيرين عشوائيين منقطعين معرفين على فضاء العينة S فى تجربة معينة من الواضح أن المجموعة الجزئية (X = x) هى حادث فى الفضاء S ، وكذلك الأمر بالنسبة لـ (Y = y) . فإذا رمزنا لهذين الحادثين بالرمزين :

$$A = [X = x], \quad B = [Y = y]$$

فإننا نستنتج باستخدام الاحتمال الشرطى للحادث B بمعلومية A الذى أوردنا تعريفه فى (١,١١) أن :

$$P (B/A) \frac{A \cap B}{P (A)} \quad : P (A) > 0$$

ومنه :

$$P [Y = y | X = x] = \frac{P [X = x , Y = y]}{P [X = x]}$$

$$= \frac{f (x, y)}{g (x)} \quad : g (x) > 0$$

ومن السهل التأكد من أن $f (x,y) \ / \ g(x)$ يحقق كل شروط توزيع الاحتمال . فلو رمزنا للطرف الأيسر فى العلاقة السابقة بالرمز $f (y/x)$ لوجدنا أن :

$$f (y/x) = \frac{f (x, y)}{g (x)} \quad : g (x) > 0$$

ونسمى الدالة الجديدة $f (y/x)$ بالتوزيع الشرطي للمتغير Y بمعلومية أن المتغير X قد أخذ القيمة x . (هذا بفرض أن كلا المتغيرين X, Y منقطعان) .

وبشكل مماثل نعرف $f(x/y)$ ليمثل التوزيع الشرطى للمتغير X بمعلومية أن المتغير Y = y بالعلاقة .

$$f (x/y) = \frac{f (x, y)}{h (y)} \quad : h (y) > 0$$

نعرف الكثافة الاحتمالية الشرطية لمتغير عشوائى مستمر X بمعلومية أن Y = y بالعلاقة :

$$f(x/y) = \frac{f(x, y)}{h(y)} \quad : h(y) > 0$$

أما بالنسبة للكثافة الاحتمالية الشرطية للمتغير المستمر Y بمعلومية أن X = x فتحدد بالعلاقة :

$$f(y/x) = \frac{f(x, y)}{g(x)} \quad : g(x) > 0$$

إذا أردنا أن نبحث عن احتمال أن يقع المتغير العشوائى المستمر X فى المجال (a, b) بمعلومية أن Y = y فعلينا أن نستخدم العلاقة :

$$P[a < X < b \mid Y = y] = \int_a^b f(x / y)\, dx$$

مثال (٢,١٤)

لنبحث فى المثال (٢,١٠) عن كل من $f(x / 2)$, $P[X = 1 \mid Y = 2]$

الحل

من الواضح أن :

$$f(x/2) = \frac{f(x, 2)}{h(2)}$$

ومن الجدول (٢,٥) نجد أن :

$$h(2) = \frac{1}{36}$$

$$f(x / 2) = 36\ f(x, 2) \quad : x = 1,2,3,4,5,6$$

لذلك فإن :

$$f(x / 2) = \begin{cases} (36) f(1,2) = (36) \left(\frac{1}{36}\right) = 1 & : x = 1 \\ (36) f(2,2) = (36) (0) = 0 & : x = 2 \\ (36) f(6,2) = (36) (0) = 0 & : x = 6 \end{cases}$$

وهكذا نجد أن جدول توزيع المتغير العشوائى X هو :

X	1	2	3	4	5	6
f(x / 2)	0	0	0	0	0	0

من الجدول نستنتج أن :

$$P[X = 1 | Y = 2] = 1$$

وهذه النتيجة تمثل احتمال أن يكون أكبر العددين اللذين ظهرا على وجهى حجرى نرد عند إلقائها هو الواحد بمعلومية أن مجموع الوجهين يساوى 2 .

مثال (٢,١٥).

لنبحث فى المثال (٢,١١) عن كل من : $f(y / x)$, $P[Y < 3 | X = \frac{1}{2}]$

الحل

وجدنا فى المثال (٢,١٣) أن :

$$g(x) = \begin{cases} \frac{3-x}{4} & : \quad 0 < x < 2 \\ 0 & : \text{فيما عدا ذلك} \end{cases}$$

ومن المعلوم أن :

$$f(y / x) = \frac{f(x, y)}{g(x)}$$

لذلك فإن :

$$F(y/x) = \begin{cases} \dfrac{(6-x-y)}{2(3-x)} & : 0 < x < 2 \\ & : 2 < y < 4 \\ 0 & : \text{فيما عدا ذلك} \end{cases}$$

ثم إن :

$$f\left(3 / \frac{1}{2}\right) = \frac{1}{2}, \ g\left(\frac{1}{2}\right) = \frac{5}{8}$$

كذلك فإن :

$$P\left[Y < 3 \mid X = \frac{1}{2}\right] = \frac{6}{10}$$

مثال (٢,١٦)

بفرض أن دالة الكثافة المشتركة للمتغيرين المستمرين X, Y هى من الشكل :

$$f(x, y) = \begin{cases} \dfrac{(1 + 5x^4)\, y^3}{8} & : 0 < x < 1 \\ & : 0 < y < 2 \\ 0 & : \text{فيما عدا ذلك} \end{cases}$$

فالمطلوب :

(١) التحقق من الشرط الثانى فى التعريف (٢,٩) بالنسبة للدالة $f(x, y)$

(٢) حساب الاحتمال $P[(X, Y) \in A]$ حيث إن :

$$A = \left\{ (x,y) \mid 0 < x < \frac{1}{2}, \ \frac{1}{4} < y < \frac{1}{2} \right\}$$

(٣) إيجاد كل من $f(x / y), h(y), g(x)$

الحل

نلاحظ أن :

أولا ـــ

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x, y)\,dx\,dy = \int_{x=0}^{1}\int_{y=0}^{2} \frac{1}{8}(1 + 5x^4)\,y^3\,dy\,dx$$

$$\int_{y=0}^{2} \left(\frac{xy^3}{8} + \frac{x^5 y^3}{8} \right)_{x=0}^{1} dy$$

$$= \int_{0}^{2} \frac{1}{4}\,y^3\,dy = 1$$

ثانيا ـــ

$$P[(X, Y) \quad A] = \int_{x=0}^{1}\int_{y=\frac{1}{4}}^{\frac{1}{2}} \frac{(1 + 5x^4)\,y^3}{8}\,dy\,dx$$

$$= \frac{(15)\,(17)}{(16)^2\,(32)^2} = 0.00097$$

ثالثا ـــ

$$g(x) = \int_{y=-\infty}^{+\infty} f(x, y)\,dy = \int_{y=0}^{2} \frac{1}{8}(1 + 5x^4)\,y^3\,dy$$

ومنه :

$$g(x) = \begin{cases} \frac{1}{2}(1 + 5x^4) & : \quad 0 < x < 1 \\ 0 & : \text{فيما عدا ذلك} \end{cases}$$

ثم أن :

$$h(y) = \int_{x=-\infty}^{+\infty} f(x,y)\,dx = \int_{x=0}^{1} \frac{1}{8}(1 + 5x^4)\,y^3\,dx$$

أى إن :

$$h(y) = \begin{cases} \frac{y^3}{4} & : 0 < y < 2 \\ 0 & : \text{ما عدا ذلك} \end{cases}$$

وأخيراً فإن :

$$f(x / y) = \frac{f(x, y)}{h(y)}$$

$$= \frac{(1 + 5x^4)}{2}$$

ومن ثم يكون :

$$f(x / y) = \begin{cases} \frac{(1 + 5x^4)}{2} & : 0 < x < 1 , 0 < y < 2 \\ 0 & : \text{فيما عدا ذلك} \end{cases}$$

والملاحظ فى هذا التمرين أن $f(x / y) = g(x)$ ، أى إن $f(x,y) = g(x)\, h(y)$ ، وهذا يقودنا إلى التعريف التالى :

تعريف (٢,١٠) المتغيران العشوائيان المستقلان

Two independence the random variables

بفرض أن X , Y متغيرين عشوائيين منقطعان أو مستمران بالكثافة الاحتمالية المشتركة $f(x,y)$ ، والكثافات الهامشية $g(x)$, $h(y)$ على الترتيب ، عندئذ نقول بأن المتغيرين العشوائيين X , Y مستقلان إحصائياً إذا وفقط إذا كان :

$$f(x,y) = g(x) \cdot h(y)$$

من أجل جميع قيم (x,y) نلاحظ أن المتغيرين العشوائيين المستمرين X,Y الواردين فى المثال (٢,١٦) السابق مستقلان إحصائيا لأنهما يحققان التعريف (٢,١٠) .

مثال (٢,١٧)

X , Y متغيران عشوائيان منقطعان لهما توزيع احتمالي مشترك معطى بالجدول التالى :

X / Y	4	10	
1	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$
3	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$
	$\frac{1}{2}$	$\frac{1}{2}$	

لنبحث فيما إذا كان هذان المتغيران مستقلين إحصائيا أم لا . من الجدول السابق نلاحظ أن جدول توزيع المتغير X هو :

X	4	10
g(x)	$\frac{1}{2}$	$\frac{1}{2}$

وأن جدول توزيع Y هو :

Y	1	3
h(x)	$\frac{1}{2}$	$\frac{1}{2}$

كما نلاحظ أنه من أجل جميع قيم (X, Y) فإن :

$$f(x,y) = g(x) \cdot h(y)$$

وهذا ما يشير إلى أن المتغيرين العشوائيين X, Y مستقلان إحصائيا .

مثال (٢,١٨)

بفرض أن التوزيع الاحتمالى المشترك للمتغيرين العشوائيين Y , X معطى بالجدول التالى :

X / Y	4	10	
1	0	$\frac{1}{2}$	$\frac{1}{2}$
3	$\frac{1}{2}$	0	$\frac{1}{2}$
	$\frac{1}{2}$	$\frac{1}{2}$	

لنبحث عن جدول توزيع كل متغير من هذين المتغيرين ، ثم لندرس استقلال هذين المتغيرين .

الحل

من الواضح من جدول التوزيع المشترك أن قيم المتغير X هى 4, 10 وأن احتمالات هذه القيم هى $\frac{1}{2}$ ، $\frac{1}{2}$ على الترتيب ، وعلى هذا الأساس فإن جدول توزيع المتغير الأول هو :

X	4	10
g(x)	$\frac{1}{2}$	$\frac{1}{2}$

كما نلاحظ بنفس الطريقة أن جدول توزيع Y هو :

Y	1	3
h(x)	$\frac{1}{2}$	$\frac{1}{2}$

ولو شكلنا الجداء $g(x) . h(y)$ ، وقارناه مع قيم $f(x,y)$ لجميع القيم (x,y) لوجدنا أن $g(x) . h(y) \neq f(x,y)$ ، وهذا يعنى أن المتغيرين X, Y ليسا مستقلين .

يمكن تعميم جميع التعاريف السابقة والمتعلقة بمتغيرين عشوائيين إلى حالة n متغير . فبفرض أن $f(x_1, \ldots, x_n)$ يمثل التوزيع الاحتمالى المشـــترك للمتغيرات العشـــوائية $X_1, \ldots X_n$ ، فإننا نلاحظ أن التوزيع الهامشى للمتغير X_1 مثلا يعطى بالعلاقة :

$$g(x_1) = \sum_{x_2} \sum_{x_3} \cdots \sum_{x_n} f(x_1, \ldots, x_n)$$

وذلك عندما تكون المتغيرات منقطعة .

وبالعلاقة :

$$g(x_1) = \int_{-\infty}^{+\infty} \cdots \int f(x_1, \ldots, x_n)\, dx_2\, dx_3 \ldots dx_n$$

فى الحالة التى تكون فيها المتغيرات السابقة مستمرة . كما نلاحظ أن التوزيع الهامشى المشترك والذى نرمز له بالرمز $\phi(x_1, x_2)$ يمكن إيجاده بإحدى العلاقتين :

١) إذا كانت المتغيرات منقطعة :

$$\phi(x_1, x_2) = \sum_{x_3} \ldots \sum_{x_n} f(x_1, \ldots, x_n)$$

٢) إذا كانت المتغيرات مستمرة :

$$= \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f(x_1, \ldots, x_n)\, dx_3\, dx_4 \ldots dx_n$$

نلاحظ أيضا أن التوزيع الشرطى المشترك joint cinditional distribution للمتغيرات X_1, X_2, X_3 بمعلومية أن $X_4 = x_4, X_5 = x_5, \ldots, X_n = x_n$ يعطى بالعلاقة :

$$f(x_1, x_2, x_3 / x_4, x_5, \ldots, x_n) = \frac{f(x_1, \ldots, x_n)}{g(x_4, \ldots, x_n)}$$

حيث يمثل $g(x_4, \dots, x_n)$ التوزيع الهامشي المشترك للمتغيرات العشوائية $X_4, X_5, \dots,$ X_n .

أما تعريف الاستقلال الإحصائي لجملة متغيرات عشوائية فسنوقه من خلال التعريف التالي :

تعريف (٢,١١) المتغير العشوائي المستقل

The independent random valiable

ليكن $(X_1, X_2, \dots, X_n)$ n متغيراً عشوائياً منقطعاً أو مستمراً بالتوزيع الاحتمالي المشترك $f(x_1, x_2, \dots, x_n)$ والتوزيعات الهامشية $f_1(x_1), f_2(x_2), \dots, f_n(x_n)$ على التتالي:

نقول بأن المتغيرات $X_1, X_2, \dots, X_n$ مستقلة إحصائيا وبالتبادل إذا وفقط إذا تحققت العلاقة :

$$f(x_1, \dots, x_n) = f_1(x_1)\, f_2(x_2) \dots f_n(x_n)$$

مثال (٢,١٩)

نفرض أن X_1, X_2, X_3 تمثل ثلاثة متغيرات عشوائية مستقلة إحصائيا وبالتبادل ، ولكل منها كثافة احتمالية معطاة بالدالة :

$$f(x) = \begin{cases} e^{-x} & : x > 0 \\ 0 & : \text{فيما عدا ذلك} \end{cases}$$

لنحسب الاحتمال :

$$P[X_1 < 3, 1 < X_2 < 2, X_3 > 2]$$

الحل

من الاستقلال التبادلي للمتغيرات X_1, X_2, X_3 وحسب التعريف (٢,١١)

نجد أن :

$$f(x_1, x_2, x_3) = f(x_1) \cdot f(x_2) \cdot f(x_3)$$

$$= e^{-x_1} \; e^{-x_2} \; e^{-x_3}$$

$$= e^{-(x_1+x_2+x_3)} \quad : x_1 > 0, x_2 > 0, x_3 > 0$$

وهكذا فإن :

$$P[X_1 < 3, 1 < X_2 < 2, X_3 > 2] =$$

$$\int_{x_1=0}^{3} \int_{x_2=0}^{2} \int_{x_3=0}^{2} e^{-(x_1+x_2+x_3)} \, dx$$

$$(1 - e^{-3})\,(e^{-1} - e^{-2})\,(1 - e^{-2}) = 0.191$$

مثال (٢,٢٠)

X, Y, Z ثلاثة متغيرات عشوائية مستمرة دالة كثافتها المشتركة من الشكل :

$$f(x, y, z) = \begin{cases} \dfrac{4xyz^2}{9} & : 0 < x < 1, 0 < y < 1, 0 < z < 3 \\ 0 & : \text{فيما عدا ذلك} \end{cases}$$

نلاحظ أن الكثافة المشتركة للمتغيرين (Y , Z) هى من الشكل :

$$\phi(y, z) = \int_{x=-\infty}^{+\infty} f(x, y, z)\, dx = \int_{x=0}^{1} \frac{4}{9}\, xyz^2\, dx$$

ومنه :

$$\phi(y, z) = \begin{cases} \frac{2}{9}\, yx^2 & : 0 < y < 1 \,,\, 0 < z < 3 \\ 0 & : \text{فيما عدا ذلك} \end{cases}$$

أما الكثافة الهامشية للمتغير Y فيمكن إيجادها من الدالة السابقة على النحو التالى :

$$h(y) = \int_{-\infty}^{+\infty} \phi(y\,,\,z)\,dz = \int_{z=0}^{3} \frac{2}{9}\, yz^2\,dz$$

$$= \frac{2}{9}\, y \left(\frac{z^3}{2} \right)_{z=0}^{3}$$

وهكذا نجد أن :

$$h(y) = \begin{cases} 2.y & : 0 < y < 1 \\ 0 & : \text{فيما عدا ذلك} \end{cases}$$

كذلك فإن الكثافة الشرطية للمتغير X بمعلومية أن $Z = 2$ ، $Y = \frac{1}{2}$ هى

$$f(x / \tfrac{1}{2}\,,\,2) = \frac{f(x, \frac{1}{2}\,,\,2)}{\phi(\frac{1}{2}\,,\,2)} = \begin{cases} 2x & : 0 < x < 1 \\ 0 & : \text{فيما عدا ذلك} \end{cases}$$

وأخيرا فإن :

$$f(x / \tfrac{1}{2}\,,\,2) = \begin{cases} 2x & : 0 < x < 1 \\ 0 & : \text{فيما عدا ذلك} \end{cases}$$

(٢,٦) التوقع الرياضى The mathematical expectation

لنلقي قطعتى نقود عشر مرات متتاليات ، ولنرمز بـ X لعدد مرات الصورة التى ظهرت فى كل إلقاء . نلاحظ أن قيم المتغير X هى 0, 1, 2 . لنفرض أنه خلال التجربة لم تظهر الصورة ثلاث مرات على التوالي ، وظهرت صورة واحدة خمس مرات ، وأخيرا ظهرت صورتين ، نلاحظ أن متوسط عدد مرات الصورة التى ظهرت فى كل إلقاء هو :

$$\frac{(0)(3) + (1)(5) + (2)(2)}{10} = (0)\left(\frac{3}{10}\right) + (1)\left(\frac{5}{10}\right) + (2)\left(\frac{2}{10}\right)$$

$$= 0.9$$

هذا العدد يمثل قيمة وسطى وليس من الضرورى أن يكون نتيجة ممكنة للتجربة . كما نلاحظ أن الكسور $\frac{3}{10}, \frac{5}{10}, \frac{2}{10}$ قد ضربت على التالى بالأعداد 0, 1, 2 وهذه الكسور تمثل التكرار النسبى للنتائج المختلفة .

لنحسب على المدى الطويل متوسط عدد مرات الصورة فى كل قذفة ، ولنرمز لهذا المتوسط والذى نسميه بالتوقع الرياضى أو القيمة المتوقعة بالرمز E (X) . من تعريف التكرار النسبى أو الاحتمال نستطيع أن نتوقع على المدى البعيد علام ظهور صورة فى حوالى ربع الوقت ، وظهور صورة واحدة فى نصف الوقت ، وظهور صورتين فى ربع الوقت الباقى لذلك فإن :

$$E(X) = (0)\left(\frac{1}{4}\right) + (1)\left(\frac{1}{2}\right) + (2)\left(\frac{1}{4}\right) = 1$$

ونتوقع أنه ظهور صورة فى المتوسط فى كل مرة عندما نلقى قطعتى نقود مرة تلو الأخرى .

من المناقشة السابقة نستنتج أن التوقع الرياضى أو المتوسط يمكن حسابه بضرب كل قيمة من قيم المتغير العشوائى بالاحتمال الموافق ، ثم بجمع هذه الجداءات . وهذا صحيح بالطبع إذا كان المتغير منقطعا . أما اذا كان المتغير مستمراً فنستبدل الجمع بالتكامل .

تعريف (٢,١٢) التوزيع الاحتمالي Probability distribution

بفرض أن $f(x)$ دالة الكثافة (أو التوزيع الاحتمالي) للمتغير العشوائي X . عندئذ نعرف التوقع الرياضي لهذا المتغير بإحدى العلاقتين :

$$E(X) = \begin{cases} \sum_{x} x f(x) & \text{إذا كان المتغير منقطعاً :} \\ \int_{-\infty}^{+\infty} x f(x)\, dx & \text{إذا كان المتغير مستمراً :} \end{cases}$$

يشترط لوجود التوقع الرياضي تحقق $\sum |x| f(x) < \infty$ أو $\int_{-\infty}^{+\infty} |x| f(x)\, dx < \infty$

مثال (٢,٢١)

لنحسب التوقع الرياضي لمتغير عشوائي منقطع X جدول توزيع من الشكل :

X	−1	0	1	2
f(x)	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{3}{8}$	$\frac{1}{4}$

نلاحظ أن توقع هذا المتغير هو :

$$E(X) = (-1)\left(\frac{1}{8}\right) + (0)\left(\frac{1}{4}\right) + (1)\left(\frac{3}{8}\right) + (2)\left(\frac{1}{4}\right)$$

ومنه :

$$E(X) = \frac{3}{4}$$

مثال (٢,٢٢)

لنبحث عن التوقع الرياضي لمتغير عشوائي Y دالة كثافته من الشكل :

$$f(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} \quad : \quad -\infty < x < +\infty$$

نلاحظ أن التوقع الرياضى لهذا المتغير هو :

$$E(Y) = \int_{-\infty}^{+\infty} y.\ f(y)\, dy = \int_{-\infty}^{+\infty} y.\ \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2}\ dy$$

$$= \frac{-1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} d\left(e^{-\frac{1}{2}y^2}\right)$$

$$= \frac{1}{\sqrt{2\pi}} \left(e^{-\frac{1}{2}y^2}\right)_{y=-\infty}^{-\infty} = 0$$

مثال (٢,٢٣)

نلاحظ أن التوقع الرياضى للمتغير X فى المثال (٢,٧) يساوى :

$$E(X) = \int_{-\infty}^{+\infty} x.\ f(x)\, dx = \int_{1}^{3} x.\ \frac{1}{2}\, dx = \frac{1}{4}\left(x^2\right)_{x=1}^{3}$$

$$= \frac{1}{4}(9-1) = 2$$

مثال (٢,٢٤)

X متغير عشوائى دالة كثافته الاحتمالية من الشكل :

$$f(x) = \begin{cases} \dfrac{4}{\pi(1+x^2)} & : 0 < x < 1 \\ 0 & : \text{فيما عدا ذلك} \end{cases}$$

لنحسب التوقع الرياضى لهذا المتغير . نلاحظ أن :

$$E(X) = \int_{-\infty}^{+\infty} x .\, f(x)\, dx = \int_0^1 \frac{4x}{\pi(1+x^2)}\, dx$$

$$= \frac{2}{\pi} \int_0^1 \frac{d(1+x^2)}{(1+x^2)} = \frac{2}{\pi} \left[\ln(1+x^2) \right]_0^1$$

$$= \frac{2}{\pi} [\ln 2 - \ln 1]$$

$$= \frac{\ln 4}{\pi}$$

مثال (٢,٢٥)

لنبحث عن التوقع الرياضي للمتغير العشوائي X الممثل لعمر نوع معين من اللمبات بالساعات ، إذا فرضنا أن الكثافة الاحتمالية لهذا العمر معطاة بالعلاقة :

$$f(x) = \begin{cases} \dfrac{375000}{x^4} & : x > 50 \\ 0 & : \text{فيما عدا ذلك} \end{cases}$$

نلاحظ أن التوقع الرياضي لعمر هذا النوع من اللمبات هو :

$$E(X) = \int_{50}^{\infty} x .\, \frac{375000}{x^4}\, dx = 75$$

ولهذا فإننا نستطيع أن نتوقع أن تخدم كل لمبة بمعدل 75 ساعة . لنفرض أن g (x) تمثل دالة ما بالمتغير العشوائى X ذى الكثافة f (x) مثلا :

$$g(x) = \begin{cases} 2x - 1 \\ x^3 \\ \sqrt{x} \\ x^2 \end{cases}$$

من الملاحظ أنه إذا أخذ المتغير X القيم 2 ,1 ,0 ,1 - ، وكان $g(x) = x^2$ فعندئذ ستكون قيم المتغير الجديد $g(x)$ هى 2 ,1 ,0 ، أما الاحتمالات الموافقة لهذه القيم فنحسبها على النحو التالى :

$$P[g(x) = 0] = P(X = 0) = f(0)$$

$$P[g(x) = 1] = P[(X = -1)] + P[(X = 1)] = f(-1) + f(1)$$

$$P[g(x) = 4] = P(X = 2) = f(2)$$

وسيكون جدول قانون توزيع $g(x)$ من الشكل :

y	0	1	4
P[g(x) = y]	f(0)	f(-1) + f(1)	f(2)

ويكون التوقع الرياضى لهذا المتغير فى هذه الحالة من الشكل :

$$E[g(x)] = \sum_y Y \cdot P[g(x) = y]$$

$$= 0. P[g(x) = 0] + 1 P[g(x) = 1] + 4 P[g(x) = 4]$$

$$= 0. f(0) + 1 [f(-1) + f(+1) + 4 f(2)$$

$$= \sum_x g(x) \cdot f(x)$$

يمكن تعميم النتيجة السابقة بالنظرية (٢,١) من أجل المتغيرات المستمرة ، والمنقطعة .

نظرية (٢,١)

بفرض X متغيراً عشوائياً بالكثافة الاحتمالية $f(x)$ ، ولتكن $g(x)$ دالة ما بهذا المتغير . عندئذ يحسب التوقع الرياضى لهذه الدالة بإحدى العلاقتين التاليتين :

إذا كان X منقطعا : $E[g(x) = \sum_x g(x) \; f(x)$

إذا كان X مستمراً :

$$= \int_{x=-\infty}^{+\infty} g(x) \quad f(x)\, dx$$

مثال (٢,٢٦)

بفرض X هو المتغير المعرف بالمثال (٢,٢١) ، وإن $g(X) = X^2$. ولنحسب التوقع الرياضى للمتغير الجديد $g(X)$.

نلاحظ فى المثال (٢,٢١) أن المتغير X هو متغير منقطع ، ولهذا فإن للمتغير الجديد $g(X)$ جدول توزيع من الشكل :

X	1	0	4
$f_{x^2}(y)$	$\frac{4}{8}$	$\frac{1}{4}$	$\frac{1}{4}$

ويكون التوقع الرياضى لـ $g(X)$ مساوياً :

$$E[g(x)] = 1 \cdot \frac{4}{8} + 0 \cdot \frac{1}{4} + 4 \cdot \frac{1}{4}$$

$$= 1.5$$

مثال (٢,٢٧)

لنحسب فى المثال (٢,١٨) كلا من $E(X)$ و $E(X^2)$ نعلم أن :

X	4	10
f(x)	$\frac{1}{2}$	$\frac{1}{2}$

ولذلك فإن :

$$E(x) = 4. \frac{1}{2} + 10 \cdot \frac{1}{2} = 7$$

كذلك فإن :

X^2	16	100
$f(x^2)$	$\frac{1}{2}$	$\frac{1}{2}$

من هذا الجدول نستنتج أن :

$$E(X^2) = (16)\left(\frac{1}{2}\right) + (100)\left(\frac{1}{2}\right) = 58$$

مثال (٢,٢٨)

نلاحظ فى المثال (٢,١٠) أن للمتغير العشوائى المنقطع X والممثل للقيمة العظمى لوجهى حجرى نرد مقذوفين ، جدول توزيع من الشكل :

X	1	2	3	4	5	6
f(x)	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{5}{36}$	$\frac{7}{36}$	$\frac{9}{36}$	$\frac{11}{36}$

لنحسب توقع هذا المتغير ، ثم توقع مربعة نلاحظ أولا أن :

$$E(X) = \sum_{x} x.\ f(x) = 1.\left(\frac{1}{36}\right) + 2.\left(\frac{3}{36}\right) + 3.\left(\frac{5}{36}\right) + 4.\left(\frac{7}{36}\right) + 5.\left(\frac{9}{36}\right) + 6.\left(\frac{11}{36}\right)$$

ومنه :

$$E(X) = 4.472$$

كما أن جدول توزيع المتغير X^2 هو :

X^2	1	4	9	16	25	36
$f(x^2)$	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{5}{36}$	$\frac{7}{36}$	$\frac{9}{36}$	$\frac{11}{36}$

والتوقع الرياضى لهذا المتغير الأخير هو :

$$E(X^2) = \sum_x x^2 . f(x)$$

$$= 1.\left(\frac{1}{36}\right) + 4.\left(\frac{3}{36}\right) + 9.\left(\frac{5}{36}\right) + 16.\left(\frac{7}{36}\right) +$$

$$25.\left(\frac{9}{36}\right) + 36.\left(\frac{11}{36}\right)$$

$$= 21.972$$

مثال (٢,٢٩)

متغير عشوائى دالة كثافته الاحتمالية من الشكل :

$$f(x) = \begin{cases} \frac{x^2}{3} & : -1 < x < 2 \\ 0 & : \text{فيما عدا ذلك} \end{cases}$$

لنبحث عن التوقع الرياضى للمتغير $g(x) = 5x - 1$

نلاحظ أن :

$$E[g(x)] = \int_{-1}^{2} (5x - 1).\frac{x^2}{3}\, dx$$

ومنه :

$$E[5x-1] = \frac{1}{3}\int_{-1}^{2}(5x^3 - x^2)\,dx = \frac{1}{3}\left[\frac{5}{4}x^4 - \frac{1}{3}x^3\right]_{-1}^{2} = \frac{21}{4}$$

سنعمم تعريف التوقع الرياضى إلى حالة متغيرين عشوائيين X , Y بالكثافة الاحتمالية المشتركة $f(x, y)$

تعريف (٢,١٣) التوقع الرياضى The mathematical expectation

بفرض أن X , Y متغيران عشوائيان بالكثافة الاحتمالية المشتركة $f(x,y)$. نعرف التوقع الرياضى للدالة $g(X, Y)$ بأحدى العلاقتين :

١) إذا كان المتغيران منقطعين

$$E[g(X, Y)] = \sum_x\sum_y g(x, y)\, f(x,y)$$

٢) إذا كان المتغيران مستمرين

$$= \int_{-\infty}^{+\infty}\int_{-\infty}^{\infty} g(x, y)\, f(x, y)\, dx\, dy$$

إن تعميم التعريف (٢,١٣) لحساب التوقع الرياضى لعدة متغيرات عشوائية ينتج مباشرة من التعريف نفسه :

مثال (٢,٣٠)

لنحسب فى المثال (٢,١٧) التوقع الرياضى للدالة $g(X, Y) = X.Y$

من التعريف (٢,١٣) نعلم أن :

$$E(X.Y) = \sum_x\sum_y x\,y\,.\,f(x, y)$$

وبالتعويض نجد أن :

$$= (4).(1).\left(\frac{1}{4}\right) + (4).(3).\left(\frac{1}{4}\right) + (10).(1).\left(\frac{1}{4}\right) + (10).(3).\left(\frac{1}{4}\right)$$

$$= 1 + 3 + \frac{5}{2} + \frac{15}{2} = 14$$

مثال (٢,٣١)

لنبحث فى المثال (٢,١٦) عن التوقع الرياضى للدالة $g(X, Y) = X^2 . Y$

الحل :

نلاحظ أن المتغيرين العشوائيين الواردين فى المثال (٢,١٦) مستمران ، لذلك فإن التوقع الرياضى للدالة $g(X, Y) = X^2 . Y$ تحسب بواسطة التعريف (٢,١٣) .

$$E(X^2 . Y) = \int_{y=0}^{2} \int_{x=0}^{1} x^2 . y \frac{(1 + 5x^4) y^3}{8} dx\ dy$$

$$= \frac{1}{8} \int_{y=0}^{2} \int_{x=0}^{1} (x^2y^4 + 5y^4x^6)\ dx\ dy$$

$$= \int_{y=0}^{2} \frac{11}{84} y^4\ dy$$

$$= \frac{88}{105}$$

مثال (٢,٣٢)

بفرض أن X, Y متغيران عشوائيان مستمران معرفان بالمثال (٢و١٦) ، ولنبحث عن $E(Y | X = x)$.

الحل :

من الواضح أن :

$$E(Y | X) = \int_{y=0}^{2} y . f(y/x)\ dy$$

$$= \int_{y=0}^{2} y . \frac{y^3}{4}\ dy$$

$$= \frac{1}{4} \int_{y=0}^{2} y^4\ dy = \frac{8}{5}$$

ملاحظة

بفرض أن $g(X, Y) = X$ ، وبالتعويض فى العلاقات الواردة فى التعريف (٢,١٣) فإننا نحصل على العلاقتين العامتين التاليتين :

١) (فى حالة متغيرات منقطعة) :

$$E(X) = \sum_{x}\sum_{y} x.\ f(x,y) = \sum_{x} x.\ g(x)$$

٢) (فى حالة متغيرات مستمرة) :

$$= \int_{-\infty}^{+\infty}\int x.\ f(x, y)\ dx\ dy = \int_{-\infty}^{+\infty} x.\ g(x)\ dx$$

حيث يمثل g(x) التوزيع الهامشى للمتغير X . من هذا نستنتج أنه لحساب التوقع الرياضى للمتغير X فى المستوى فإن بإمكاننا استخدام التوزيع المشترك للمتغيرين X, Y ، أو التوزع الهامشى للمتغير X . وبنفس الطريقة نعرف :

١) (فى حالة متغيرات منقطعة) :

$$E(Y) = \sum_{x}\sum_{y} y.\ f(x,y) = \sum_{y} y.\ h(y)$$

٢) (فى حالة متغيرات مستمرة) :

$$\int_{-\infty}^{+\infty}\int y.\ f(x, y)\ dx\ dy = \int_{-\infty}^{+\infty} y.\ h(y)\ dy$$

حيث يمثل h(y) التوزيع الهامشى للمتغير Y .

(٢,٧) قوانين التوقع الرياضى Laws of Expectation

سنبرهن الآن بعض القوانين المفيدة فى حساب التوقع الرياضى ، وهذه الدساتير أو النظريات ستسمح لنا بحساب التوقعات الرياضية بوساطة بعض التوقعات المعروفة ، وستكون جميع النتائج صحيحة من أجل المتغيرات المنقطعة والمستمرة على حد سواء ، وسنقوم ببرهان هذه النتائج من أجل متغيرات مستمرة .

نظرية (٢,٢)

إذا كان b. a ثابتين فإن :

$$E(aX + b) = a.\ E(X) + b$$

البرهان :

من تعريف التوقع الرياضى نجد أن :

$$E\ (aX + b) = \int_{-\infty}^{+\infty} (ax + b).\ f(x)\ dx$$

$$= a. \int_{-\infty}^{+\infty} x.\ f(x)\ dx + b. \int_{-\infty}^{+\infty} f(x)\ dx$$

والتكامل الثانى فى الطرف الأيمن ما هو إلا جُراء ضرب a فى التوقع الرياضى لـ X ، أما التكامل الأول فيساوى الواحد . لذلك فإن :

$$E(aX + b) = a.\ E(x) + b$$

نتيجة (٢,١)

بفرض $a = 0$ ، فإننا نلاحظ أن $E(b) = b$
أى إن التوقع الرياضى لثابت يساوى الثابت نفسه .

نتيجة (٢,٢)

بفرض $b = 0$ ، فإننا نلاحظ أن $E(a.X) = a.E(X)$
أى إن التوقع الرياضى لجُراء ضرب ثابت بمتغير عشوائى يساوى إلى حاصل ضرب الثابت بالتوقع الرياضى لذلك المتغير .

نظرية (٢,٣)

إن التوقع الرياضى لمجموع أو الفرق بين دالتين (بالمتغير العشوائى X) يساوى مجموع أو الفرق بين فضل التوقع الرياضى لهاتين الدالتين :

$$E[g(X) \pm h(X)] = E[g(X)] \pm E[h(X)]$$

البرهان

من تعريف التوقع الرياضى نجد أن :

$$E[g(X) \pm h(X)] = \int_{-\infty}^{+\infty} [g(x) \pm h(x)] . f(x)\, dx$$

$$= \int_{-\infty}^{+\infty} g(x) . h(x)\, dx \pm \int_{-\infty}^{+\infty} h(x) f(x)\, dx$$

$$= E[g(X)] \pm E[h(X)]$$

مثال (٢,٣٣)

بفرض X هو المتغير العشوائى الوارد فى المثال (٢,١٦) ، احسب التوقع الرياضى للدالة $(X + 1)^2$.

الحل

نعلم أن دالة الكثافة X فى المثال (٢,١٦) هى من الشكل :

$$g(x) = \begin{cases} \frac{1}{2}(1 + 5x^4) & : 0 < x < 1 \\ 0 & : \text{فيما عدا ذلك} \end{cases}$$

ومن الواضح أن :

$$E(X+1)^2 = E(X^2 + 2X + 1)$$

$$= E(X^2) + 2E(X) + 1$$

كما نجد :

$$E(X) = \int_0^1 x . \frac{1}{2}(1 + 5x^4)\, dx = \frac{2}{3}$$

$$E(X^2) = \int_0^1 x^2 . \frac{1}{2}(1 + 5x^4)\, dx = \frac{11}{21}$$

وبالتعويض نجد أن :

$$E(X+1)^2 = \frac{11}{21} + 2 . \frac{2}{3} + 1$$

$$= \frac{60}{21}$$

مثال (٢,٣٤)

وجدنا فى المثال (٢,٢١) أن $E(X) = \frac{3}{4}$ ، لنفتش عن $E(2x+1)$ من الواضح أن :

$$E(2X+1) = 2 . E(X) + 1$$

وبالتعويض نجد أن :

$$= 2 . \frac{3}{4} + 1 = \frac{5}{2}$$

لنفرض الآن أن لدينا متغيرين عشوائيين Y, X بالكثافة الاحتمالية المشتركة $f(x,y)$ ، ولندرس نظرية تتعلق بمجموع وفرق دالتين بهذين المتغيرين .

نظرية (٢,٤)

إن التوقع الرياضى لمجموع أو فرق دالتين أو أكثر بهذين المتغيرين يساوى إلى مجموع أو فرق التوقعين الرياضيين هاتين الدالتين .

$$E[g(X,Y) \pm h(X,Y)] = E[g(X,Y)] \pm E[h(X,Y)]$$

البرهان

من التعريف (٢,١٣) نجد أن :

$$E[g(X,Y) \pm h(X,Y)] = \int_{-\infty}^{+\infty}\int [g(x,y) \pm h(x,y)]\, f(x,y)\, dx\, dy$$

$$= \int_{-\infty}^{+\infty}\int h(x,y) . f(x,y)\, dx\, dy \quad \pm \int_{-\infty}^{+\infty}\int g(x,y)\, f(x,y)\, dx\, dy$$

$$= E[g(X,Y)] \pm E[h(X,Y)]$$

نتيجة (٢,٣)

بفرض أن $h(X, Y) = Y, g(X, Y) = X$ فإننا نجد أن :

$$E(X \pm Y) = E(X) \pm E(Y)$$

وهذه النتيجة الأخيرة هامة جدا ، فهى تقرر أن التوقع الرياضى لمجموع (أو فرق) متغيرين عشوائيين ماهو إلا مجموع (أو فرق) التوقعين الرياضيين لهذين المتغيرين على الترتيب .

نظرية (٢,٥)

إذا كان المتغيران العشوائيان X, Y مستقلين ، فإن :

$$E(X . Y) = E(X) . E(Y)$$

البرهان

من تعريف التوقع الرياضى نجد أن :

$$E(X. Y) = \int_{-\infty}^{+\infty}\int x. y. f(x, y)\, dx.\, dy$$

وبما أن X, Y مستقلان فيمكننا أن نكتب :

$$f(x, y) = g(x) . h(y)$$

حيث يمثل كلا من $h(y), g(x)$ التوزعات الهامشية لكل من Y, X على الترتيب ، ولذلك فإن :

$$E(X.Y) = \int\int_{-\infty}^{+\infty} x.\, y.\, g(x)\,.\, h(y)\, dx\,.\, dy$$
$$= \int_{-\infty}^{+\infty} x.\, g(x).\, dx \int_{-\infty}^{+\infty} y.\, h(y)\, dy$$

$$= E(X)\,.\,E(Y).$$

مثال (٢,٣٥)

لنحسب التوقع الرياضى لحاصل ضرب المتغيرين العشوائيين المستمرين الواردين فى المثال (٢,١٦) .

الحل

لاحظنا فى المثال المذكور أن المتغيرين X , Y يمثلان متغيرين عشوائيين مستقلين ، وأن دالة كل منهما هى من الشكل :

$$h(y) = \begin{cases} \dfrac{y^3}{4} & : 0 < y < 2 \\ 0 & : \text{فيما عدا ذلك} \end{cases}$$

$$g(x) = \begin{cases} \dfrac{1}{2}(1 + 5x^4) & : 0 < x < 1 \\ 0 & : \text{فيما عدا ذلك} \end{cases}$$

كذلك حسبنا فى المثال (٢,٣٣) التوقع الرياضى للمتغير X ، ووجدنا أن :

$$E(X) = \frac{2}{3}$$

لنحسب الآن توقع المتغير Y

$$E(Y) = \int_0^2 y \cdot \frac{y^3}{4}\, dy = \frac{1}{4}\int_0^2 y^4\, dy$$

$$= \frac{8}{5}$$

وبالتعويض فى النظرية (٢,٥) نجد أن :

$$E(X) \cdot E(Y) = \frac{2}{3} \cdot \frac{8}{5} = \frac{16}{15}$$

لنحسب الآن دون استخدام فكرة استقلال التوقع الرياضى لجراء الضرب .

$$E(X \cdot Y) = \int_0^2 \int_0^1 x \cdot y\, \frac{1}{8}\, (1 + 5x^4)\, y^3\, dy\, dx$$

$$= \int_0^1 \frac{x}{8}\, (1 + 5x^4)\, dx \cdot \int_0^2 y^4\, dy$$

$$= \frac{4}{5} \int_0^1 (x + 5x^5)\, dx = \frac{4}{5} \left(\frac{1}{2} + \frac{5}{6} \right) = \frac{16}{15}$$

وهى نفس النتيجة التى وجدناها سابقا .

(٢,٨) التوقعات الرياضية الخاصة (التباين ـــ التغير)

Special mathematical expectation

تقودنا النظرية (٢,١) إلى قيمة متوقعة تدعى بالعزم من المرتبة K حول المبدأ للمتغير العشوائى X فيما لو عوضنا $g(x) = X^k$. نرمز لهذا العزم بالرمز μ لذلك فإن .

إذا كان المتغير منقطعاً : $\mu'_k = E|(X)^k| = \sum_x x^k \cdot f(x)$

إذا كان المتغير مستمراً : $= \int x^k \cdot f(x)\, dx$

نلاحظ أنه إذا كان $K = 0$ فان : $\mu'_0 = E(X^0) = 1$ ،

ومنه سواء كان X مستمراً أو منقطعاً فإن : $E(1) = 1$.

أما إذا كان K = 1 فإننا نجد أن $\mu'_1 = E(X)$. سنرمز لهذا العزم بالرمز μ_x ، أو بالرمز μ . ولذلك فإن :

$$\mu = \mu'_1 = E(X)$$

أما إذا عوضنا فى النظرية (٢,١) عن $g(x) = (X - \mu)^k$ فعندئذ تأخذ هذه النظرية الشكل التالى :

إذا كان X منقطعاً :
$$E(X - \mu)^k = \sum_x (x - \mu) f(x)$$

إذا كان X مستمرا :
$$= \int_{-\infty}^{+\infty} (x - \mu)^k . f(x)\, dx$$

نرمز لهذا العزم الجديد بالرمز μ_k ، ويدعى بالعزم من المرتبة K حول المتوسط . إن العزم حول المتوسط من المرتبة الثانية والذى نرمز له بالرمز μ_2 يعطينا معلومات عن تغير القياس حول متوسطه . سنسمى هذا العزم بتباين المتغير العشوائى X وسنرمز له بالرمز σ^2 ، وهكذا نجد أن :

$$\sigma^2 = \mu_2 = E(X - \mu)^2$$

نسمى الجذر التربيعى الموجب للتباين بالانحراف المعيارى للمتغير العشوائى .

نظرية (٢,٦)

إن تباين أى متغير عشوائى X يعطى بالعلاقة :

$$\sigma^2 = = E(X^2) - \mu^2$$

البرهان

من تعريف العزم من المرتبة الثانية حول المتوسط نجد أن :

$$\sigma^2 = E[(x - \mu)^2]$$

$$= E(X^2 + \mu^2 - 2\mu X)$$

$$= E(X^2) + E(\mu^2) - 2\mu E(X)$$

ومن المعلوم أن :

$$E(\mu^2) = \mu^2 \quad , \quad E(X) = \mu$$

لذلك فإن :

$$\sigma^2 = E(X^2) - \mu^2$$

مثال (٢,٣٦)

بفرض X متغير منقطع مُعرَّف بالجدول :

X	4	10
f(x)	$\frac{1}{2}$	$\frac{1}{2}$

وجدنا فى المثال (٢,٢٧) أن :

$$E(x) = 7 , E(X^2) = 58$$

وحسب النظرية (٢,٦) فإن تباين هذا المتغير يساوى :

$$\sigma^2 = E(X^2) - (E(x))^2$$

$$\sigma^2 = 58 - 49$$

$$\sigma^2 = 9$$

وكذلك فإن الانحراف المعيارى لهذا المتغير هو :

$$\sigma = 3$$

مثال (٢,٣٧)

وجدنا فى المثال (٢,٢٣) أن للمتغير العشوائى المستمر X بالكثافة :

$$g(x) = \begin{cases} \frac{1}{2}(1 + 5x^4) & : 0 < x < 1 \\ 0 & \text{: فيما عدا ذلك} \end{cases}$$

توقعا رياضياً $E(X) = \frac{2}{3}$ وعزماً من المرتبة الثانية حول المبدأ يساوى $E(X^2) = \frac{11}{12}$ وحسب النظرية (٢,٦) فإن تباين هذا المتغير يساوى :

$$\sigma^2 = E(X^2) - (E(X))^2 = \frac{11}{21} - \frac{4}{9} = \frac{5}{63}$$

أما انحرافه المعيارى الـ ـرى إلى جذر تباينه فهو :

$$\sigma = \frac{\sqrt{5}}{\sqrt{63}}$$

إذا عوضنا فى التعريف (٢,١٣) عن الدالة :

$$g(x, y) = (x - \mu_x).(Y - \mu_y)$$

حيث يمثل :

$$\mu_y = E(Y) , \mu_x = E(X)$$

فإن هذا التعريف يقودنا عندئذ إلى قيمة متوقعة جديدة تدعى بتمام التباين (Covariance) للمتغيرين X, Y والذي نرمز له بأحد الرمزين Cov (X,Y) أو σ_{xy} ، لذلك فإن :

$$\sigma_{xy} = E[(X - \mu_x).(Y - \mu_y)]$$

١) إذا كان Y, X منقطعين :

$$= \sum_x \sum_y (x - \mu_x).(y - \mu_y).f(x,y)$$

٢) إذا كان Y, X مستمرين :

$$= \int_{-\infty}^{+\infty}\int (x - \mu_x).(y - \mu_y) f(x, y)\, dx\, dy$$

نظرية (٢,٧)

إن تمام تباين متغيرين عشوائيين Y, X بالمتوسطين μ_x , μ_y على الترتيب يعطى بالعلاقة التالية :

$$\sigma_{xy} = E(X.Y) - \mu_x \cdot \mu_y$$

نلاحظ أن :

$$\sigma_{xy} = E[(X - \mu_x)(Y - \mu_y)]$$
$$= E[X.Y - X.\mu_y - Y.\mu_y + \mu_x \cdot \mu_y]$$

باستخدام النظرية (٢,٢) والنتيجة (٢,٣) نجد أن :

$$\sigma_{xy} = E(X.Y) - E(X.\mu_y) - E(Y.\mu_x) + \mu_x \cdot \mu_y$$

وحيث إن :

$$\mu_y = E(Y), \mu_x = E(X)$$

فإننا نجد أن :

$$\sigma_{XY} = E(X.Y) - \mu_Y \cdot \mu_X - \mu_X \cdot \mu_Y + \mu_X \cdot \mu_Y$$
$$= E(X.Y) - \mu_X \cdot \mu_Y$$

مثال (٢,٣٨)

Y, X متغيران عشوائيان مستمران بالكثافة المشتركة :

$$f(x,y) = \begin{cases} 4xy & : 0 < x < 1 \\ & \;\; 0 < y < 1 \\ 0 & : \text{فيما عدا ذلك} \end{cases}$$

لنبحث عن تمام المتغيرين X, Y ، نلاحظ أن :

$$\mu_x = E(X) = \int_{x=0}^{1} x \left(\int_{x=0}^{1} 4x\, y\, dy\right) dx = \frac{2}{3}$$

$$\mu_y = E(Y) = \int_{0}^{1} y \left(\int_{0}^{1} 4x\, y\, dx\right) dy = \frac{2}{3}$$

كما نلاحظ أيضاً أن :

$$E(X.Y) = \int_{y=0}^{1} \int_{x=0}^{1} 4x^2 y^2 \, dx \, dy = \frac{4}{9}$$

وأخيراً فإن :

$$\sigma_{xy} = E(X.Y) - \mu_x . \mu_y$$

$$= \frac{4}{9} - \left(\frac{2}{3}\right) . \left(\frac{2}{3}\right)$$

$$= 0$$

(٢,٩) خواص التباين Properties of the variance

فيما يلى سنبرهن أربع نظريات تعتبر مفيدة فى حساب التابين أو الانحراف المعيارى . سنرمز ب $\sigma^2_{g(x)}$ و $\mu_{g(x)}$ للتباين ، ومتوسط الدالة g(x) على التتالى .

نظرية (٢,٨)

بفرض X متغير عشوائى بالتوزيع الاحتمالى f(x) . عندئذ يكون تباين الدالة g(x) يعطى بالعلاقة :

$$\sigma^2_{g(x)} = [\, E\{g(x) - \mu_{g(x)}\}^2 \,]$$

نلاحظ أن برهان هذه النظرية ينتج مباشرة من تعريف تباين متغير عشوائى وعلى اعتبار أن g(x) هى دالة بالمتغير العشوائى X فهو من جديد متغير عشوائى .

نظرية (٢,٩)

بفرض X متغير عشوائى و b ثابتاً عندئذ يكون :

$$\sigma^2_{X+b} = \sigma^2_X = \sigma^2$$

نلاحظ من النظرية السابقة ، وبفرض $g(X) = X + b$ أن :

$$\sigma^2_{x+b} = E[\{(X+b) - \mu_{X+b}\}^2]$$

ونعلم أن :

$$\mu_{x+b} = E(X+b) = \mu + b$$

إذا :

$$\sigma^2_{X+b} = E[\{(X+b) - (\mu + b)\}^2]$$

$$= E(X-\mu)^2 = \sigma^2$$

وتفسير هذه النظرية أن تباين أى متغير عشوائى لا يتأثر بإضافة أو طرح عدد ثابت إلى هذا المتغير .

نظرية (٢,١٠)

إذا كان X متغيرا عشوائيا و a ثابتا ما ، فعندئذ يكون :

$$\sigma^2_{aX} = a^2\sigma^2_X = a^2 \cdot \sigma^2$$

البرهان

حسب تعريف التباين نجد أن :

$$\sigma^2{}_{a(x)} = E\{(aX - \mu_{g(x)})^2\}$$

وحيث إن :

$$\mu_{a(x)} = E(a.X) = a.E(X) = a.\mu$$

لذلك فإن :

$$\sigma^2_{aX} = E\{[a(X-\mu)]^2\}$$

$$= a^2 \ E(X-\mu)^2$$

$$= a^2 \ \sigma^2$$

وهذا يعنى أنه إذا ضرب متغير بثابت أو قسم على ثابت فإن التباين يجب أن يضرب بمربع الثابت أو يقسم على مربع الثابت على الترتيب .

نظرية (٢,١١)

بفرض أن X, Y متغيرات عشوائيات بالكثافة الاحتمالية المشتركة g(x,y) عندئذ يكون :

$$\alpha^2_{aX+bY} = a^2\,\sigma^2_X + b^2\,\sigma^2_Y + 2\,a.b.\sigma_{X.Y}$$

البرهان

من تعريف التباين نجد أن :

$$\sigma_{aX+bY} = E\,[\,\{\,(aX + bY) - E(aX + bY)\,\}^2\,]$$

غير أن :

$$E(aX + bY) = a\,\mu_X + b\,\mu_Y$$

ومنه :

$$\sigma^2_{aX+by} = E\,[\,\{\,(a\,X + b\,y) - (a\mu_X + b\,\mu_Y)\,\}^2\,]$$

$$= E\,[\,\{\,a\,(x - \mu_X) + b\,(Y - \mu_Y)\,\}^2\,]$$

$$= a^2E(x - \mu_X)^2 + b^2E(Y - \mu_Y)^2 + 2a.bE(X - \mu_X)\,.\,(Y - \mu_Y)$$

$$= a^2\,\sigma^2_X + b^2\,\sigma^2_Y + 2a.b\,\sigma_{ZY}$$

نتيجة (٢,٤)

إذا كان المتغيران Y, X مستقلين فعندئذ يكون :

$$\sigma^2_{aX+bY} = a^2\,.\,\sigma^2_X + b^2\,.\,\sigma^2_Y\,.$$

بالحقيقة من تعريف تمام التباين σ_{xy} ، والنظرية (٢,٧) نجد ان :

$$\sigma_{XY} = E(X\,.\,Y) - \mu_X\,\mu_Y$$

ومن النظرية (٢,٥) ، وحيث إن المتغيرين مستقلان لذا فإن :

$$\sigma_{XY} = E(X)\,.\,E(Y) - \mu_X\,.\,\mu_Y = 0$$

وبحسب النظرية (٢,١١) نجد أن :

$$\sigma^2_{aX+bY} = a^2\,.\,\sigma^2_X + b^2\,\sigma^2_Y + 2a.b.0$$

وأخيرا فإننا نجد :

$$\sigma^2_{aX + bY} = a^2 \sigma^2_x + b^2 \sigma^2_Y$$

نتيجة (٢,٥)

إذا كان Y, X مستقلين فعندئذ :

$$\sigma^2_{aX + bY} = a^2 . \sigma^2_X + b^2 \sigma^2_Y$$

هذه النتيجة تنتج مباشرة بوضع b - عوضا عن b فى النتيجة (٢,٤) .

مثال (٢,٣٩)

بفرض أن تباين المتغيرين Y, X هما على الترتيب $\sigma^2_X = 2$, $\sigma^2_Y = 4$ وبفرض أن تمام تباينها هو $\sigma_{XY} = -2$ ، فأوجد تباين المتغير $Z = 5X + 3Y - 4$.

الحل :

بحسب النظريتين (٢,٩) ، (٢,١١)

$$\sigma^2_{5x+3y-4} = \sigma^2_{5x+3y}$$

وبحسب الفرض نجد :

$$\sigma^2_z + 25\,\sigma^2_x + 9\,\sigma^2_y + 2.3.5\,\sigma_{x.y}$$

$$\sigma^2_Z = (25)(2) + (9)(4) + 2.3.5.(-2)$$

$$= 50 + 36 - 60$$

$$= 26$$

(٢,١٠) نظرية تشيبيشيف Chebyshev's theorem

إن الدور الهام الذى يلعبه الانحراف المعيارى لمتغير عشوائى كمقياس لتباين قيم المتغير العشوائي عن وسطه (قيمة المتوقعة) توضحه لنا متباينة تشبيشيف الآتية :

نظرية تشبيشيف

إن احتمال أن يقع متغير عشوائى X فى المجال $(\mu - k\sigma, \mu + k\sigma)$ أكبر أو يساوى $\left(1 - \frac{1}{k^2}\right)$ ، حيث يمثل K ثابتاً ما ، σ الانحراف المعيارى للمتغير X, μ وسطه .

$$P[\mu - K\sigma < X < \mu + K\sigma] \geq 1 - \frac{1}{k^2}$$

البرهان

من تعريف تباين متغير عشوائى X نجد أن :

$$\sigma^2 = E[(x - \mu)]^2 = \int_{-\infty}^{+\infty} (x - \mu)^2 . f(x)dx$$

$$= \int_{-\infty}^{\mu - k\sigma} (x - \mu)^2 . f(x)dx + \int_{\mu - k\sigma}^{\mu + k\sigma} (x - \mu)^2 . f(x)\,dx + \int_{\mu + k\sigma}^{+\infty} (x - \mu)^2 f(x)\,dx.$$

ومنه :

$$\sigma^2 \geq \int_{-\infty}^{\mu - k\sigma} (x - \mu)^2 f(x)\,dx + \int_{\mu + k\sigma}^{+\infty} (x - \mu)^2 f(x)\,dx$$

على اعتبار أن التكامل الأوسط ليس سالباً وبما أن $|x - \mu| > K\sigma$ فى الحالتين أى أن

$$(x - \mu)^2 \geq k^2\sigma^2 \quad , \quad X \leq \mu - k\sigma \quad , \quad X \geq \mu + k\sigma$$

فهذا يعنى أن :

$$\sigma^2 \geq \int_{-\infty}^{\mu - k\sigma} k^2\sigma^2 f(x)\,dx + \int_{\mu + k\sigma}^{+\infty} k^2 . \sigma^2 f(x)\,dx$$

ومنه :

$$\frac{1}{k^2} \geq \int_{-\infty}^{\mu - k\sigma} f(x)dx + \int_{\mu + k\sigma}^{+\infty} f(x)dx$$

وأخيراً فإننا نستنتج أن :

$$P[\mu - k\sigma < X < \mu + K\sigma] = \int_{\mu - k\sigma}^{\mu + k} f(x)\,dx \geq 1 - \frac{1}{k^2}$$

نلاحظ أنه في الحالة التي يكون فيها K = 2 فإن المتغير X احتمالا على الأقل $\frac{3}{4} = \frac{1}{4} - 1$ للوقوع فى المجال $(\mu - 2\sigma, \mu + 2\sigma)$. وهذا يعنى أن ثلاث أرباع الملاحظات من أى توزيع ستقع فى المجال $\mu \pm 2\sigma$. وبشكل مشابه نجد أن ثمانية أتساع الملاحظات $\frac{8}{9}$ أو أكثر من أى توزيع ستقع فى المجال $\mu \pm 3\sigma$.

مثال (٢,٤٠)

متغير عشوائى X توقعه $\mu = 8$ وتباينه $\sigma_2 = 9$ ، وتوزيعه غير معرف . لنبحث عن

1. $p[-4 < X < 20]$
2. $P[|x - 8| \geq 6]$

الحل

نلاحظ أن :

$$P[-4 < X < 20] = P[8 - 4.3 < X < 8 + 4.3]$$

$$\geq 1 - \frac{1}{4^2} = \frac{15}{16}$$

$$\geq \frac{15}{16}$$

كذلك فإن :

$$P[|X - 8| \geq 6] = 1 - P[|x - 8| < 6]$$
$$= 1 - P[8 - 6 < X < 8 + 6]$$
$$= 1 - P[8 - 2.3 < X < 8 + 2.3]$$
$$\leq 1 - (1 - \frac{1}{2^2})$$

$$\leq \frac{1}{4}$$

ملاحظة :

إن النتائج التى تقدمها متباينة تشبيشيف تكون عادة ضعيفة لأننا نعلم أن احتمال أن يقع المتغير العشوائى X فى المجال $(\mu - 2.\sigma\ \mu + 2\sigma)$ ليس أقل من $\frac{3}{4}$ ، ولكن لا نعلم على وجه الدقة إلى أى حد يمكن أن يكون هذا الاحتمال أكبر من $\frac{3}{4}$. أما عندما يكون التوزيع الاحتمالى للمتغير X معلوماً بالنسبة لنا فعندئذ يمكن أن نحدد بالضبط الاحتمال السابق .

تمارين محلولة

تمرين (١)

وضعت نقطة من اللبان على أحد أوجه قطعة نقود بحيث أصبح $P(H) = \frac{1}{4}$ و $P(T) = \frac{3}{4}$ ، ثم ألقيت هذه القطعة ثلاث مرات . أوجد جدول قانون توزيع المتغير X الممثل لعدد الصور التى ظهرت .

الحل

من الواضح أن فضاء العينة فى هذه التجربة هو الفضاء

$$S = [(HHH), (HTT), (HHT), (TTH), (HTH), (TTT), (THH), (THT)]$$

ونلاحظ أن قيم المتغير X المعرف على الفضاء هى 0, 1, 2, 3 كما نلاحظ أيضا أن :

$$P(HHH) = \frac{1}{4} \cdot \frac{1}{4} \cdot \frac{1}{4} = \frac{1}{64}$$

$$P(HTT) = \frac{1}{4} \cdot \frac{3}{4} \cdot \frac{3}{4} = \frac{9}{64}$$

$$P(HHT) = \frac{1}{4} \cdot \frac{1}{4} \cdot \frac{3}{4} = \frac{3}{64}, \quad P(THT) = \frac{3}{4} \cdot \frac{1}{4} \cdot \frac{3}{4} = \frac{9}{64}$$

$$P(TTH) = \frac{9}{64}, \quad P(TTT) = \frac{3}{4} \cdot \frac{3}{4} \cdot \frac{3}{4} = \frac{27}{64}$$

$$P(HTH) = \frac{3}{64}, \quad P(THH) = \frac{3}{4} \cdot \frac{1}{4} \cdot \frac{1}{4} = \frac{3}{64}$$

ثم أن :

$$f(0) = P(TTT) = \frac{27}{64}$$

$$f(1) = P(HTT) + P(TTH) + P(THT) = \frac{27}{64}$$

$$f(2) = P(HHT) + P(THH) + P(HTH) = \frac{9}{64}$$

$$f(3) = P(HHH) = \frac{1}{64}$$

وهكذا نجد أن :

X	0	1	2	3	
f(x)	$\frac{27}{64}$	$\frac{27}{64}$	$\frac{9}{64}$	$\frac{1}{64}$	$\sum_x f(x) = 1$

والملاحظ أن :

$$F(x) = \begin{cases} 0 & : x < 0 \\ \frac{27}{64} & : 0 \leq x < 1 \\ \frac{56}{64} & : 1 \leq x < 2 \\ \frac{63}{64} & : 2 \leq x < 3 \\ 1 & : x > 3 \end{cases}$$

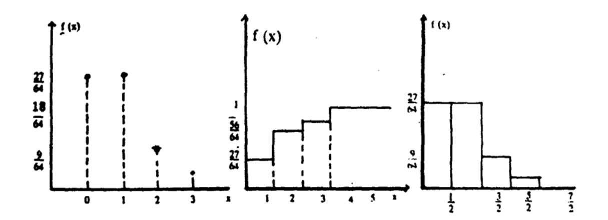
من أجل : $x < 0$

من أجل : $0 \leq x < 1$

من أجل : $1 \leq x < 2$

من أجل : $2 \leq x < 3$

أما بالنسبة للمنحنيات الممثلة لدالة التوزيع الاحتمالى والتوزيع التراكمى والمضلع التكرارى فهى :



تمرين (٢)

بفرض أن X يمثل ضعف الرقم الذى ظهر لدى إلقاء حجر نرد . ما هو جدول قانون توزيع المتغير X ، ارسم المنحنيات البيانية الموافقة ؟

الحل

نلاحظ أن عناصر فضاء العينة لتجربة إلقاء حجر النرد مرة واحدة هى :

$$S = \{1, 2, 3, 4, 5, 6\}$$

إذن فقيم المتغير X هى :

$$X : 2, 4, 6, 8, 10, 12$$

أما الاحتمالات الموافقة فهى :

$$P(X = 2) = \frac{1}{6},\ P(X = 4) = \frac{1}{6}\quad, \ldots,\ P(X = 12) = \frac{1}{6}$$

وجدول قانون التوزيع هو :

x	2	4	6	8	10	12	
f (x)	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\Sigma f(x) = 1$

كما نلاحظ أن :

$$F(2) = \frac{1}{6}$$

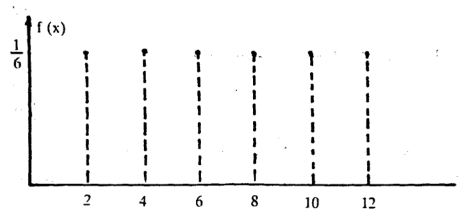
$$F(4) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$$

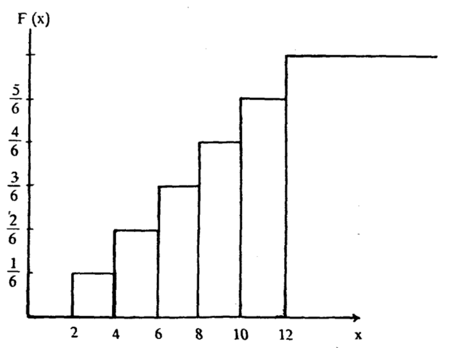
$$F(6) = \frac{1}{6}\ \frac{1}{6} + \frac{1}{6} = \frac{1}{2}$$

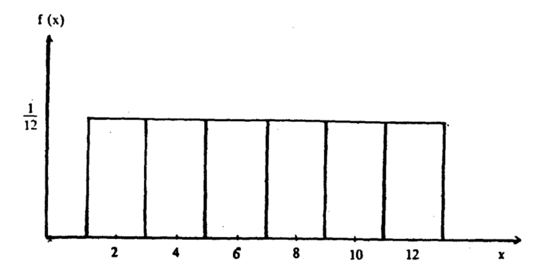
$$F(8) = \frac{4}{6} = \frac{2}{3}$$

$$F(10) = \frac{5}{6}$$

$$F(12) = 1$$


f (x)
$\frac{1}{6}$
2
4
6
8
10
12


F (x)
$\frac{5}{6}$
$\frac{4}{6}$
$\frac{3}{6}$
$\frac{2}{6}$
$\frac{1}{6}$
2
4
6
8
10
12
x


f (x)
$\frac{1}{12}$
2
4
6
8
10
12
x

تمرين (٣)

في المثال (٢,٨) أوجد $F(x)$ ، ثم أحسب بواسطته $P(2.5 \leq x < 4)$.

الحل

من المعلوم أن للمتغير المستمر كثافة احتمالية فوق المجال (2,5) وفيما عدا ذلك فإن الكثافة معدومة ، وعلى هذا الأساس . نلاحظ أنه إذا كان $x \leq 2$ فإن $F(x) = 0$ أما إذا كان $2 < x < 5$ فعندئذ :

$$f(x) = \int_{-\infty}^{x} f(u)\,du = \int_{-\infty}^{2} f(u)\,du + \int_{2}^{x} f(u)\,du$$

$$= \int_{-\infty}^{2} 0.du + \int_{2}^{x} \frac{2(1+u)}{27}\,du = \frac{1}{27}(x^2 + 2x - 8)$$

أما إذا كان $x > 5$ فإن :

ولذلك فإن :

$$F(x) = \begin{cases} 0 & \text{if} \quad x \leq 2 \\ 1 & \text{if} \quad x \geq 5 \\ \frac{1}{27}(x^2 + 2x - 8) & \text{if} \quad 2 < x < 5 \end{cases}$$

تمرين (٤)

بفرض أن X يمثل مجموع الوجهين اللذين ظهرا لدى إلقاء حجري نرد متماثلين ومتوازيين ، فتش عن التوزيع الاحتمالي للمتغير X .

الحل

الملاحظ أن قيم المتغير العشوائي المنقطع X ، والممثل لمجموع الوجهين اللذين ظهرا هي :

X: 2, 3, 4, 5, ... , 10, 11, 12

أما الاحتمالات الموافقة لهذه القيم فنحددها بواسطة العلاقة : $f(x) = P[X = x]$. فمثلا من أجل $x = 2$ نجد أن :

$$f(2) = P[X = 2] = P[2 = \text{مجموع الوجهين}] = P[1, 1] = \frac{1}{36}$$

كذلك فإن f(3) يحسب بنفس الطريقة :

$$f(3) = P[X = 3] = P[3 = \text{مجموع الوجهين}] = P[(1,2), (2,1)] = \frac{2}{36}$$

وبنفس الطريق نجد بقية قيم f(x) . وبترتيب هذه القيم والاحتمالات في جدول نجد أن :

x	2	3	4	5	6	7	8	9	10	11	12
f(x)	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

تحقق من أن $\sum_x f(x) = 1$ ثم مثل التوزع السابق بيانيا .

تمرين (٥)

اشترى سعيد بطاقتين من أصل خمسة آلاف بطاقة من يانصيب خيري بريال واحد للبطاقة الواحد . ما هو توزيع ربح سعيد في هذه الحالة ؟

الحل

إذا رمزنا لربح سعيد بالرمز X ، واعتبرنا أن خسارته هى ربح قدره (2 -) ريالا فإننا نجد أن للمتغير X قيمتين فقط هما 2,4998 - والاحتمالات الموافقة لهذه القيم هى :

$$f(-2) = P[X = -2] = \frac{4998}{5000} = 0.9996$$

$$f(4998) = P[X = 4998] = \frac{2}{5000} = 0.0004$$

وهكذا نجد أن جدول قانون توزيع ربح بسعيد فى هذه العملية هو :

X	−2	2998
f(x)	0.9996	0.0004

لاحظ أن $\sum_x f(x) = 1$

تمرين (٦)

بفرض X متغير عشوائى مستمر دالة كثافته من الشكل :

$$f(x) = \begin{cases} \frac{1}{2}x & : 0 \le x \le 2 \\ 0 & : \text{فيما عدا ذلك} \end{cases}$$

فالمطلوب تحديد دالة التوزيع F(x) لهذا المتغير ، ثم رسم المنحنيين البيانيين لكل من F(x) و f(x) .

الحل

نلاحظ أن :

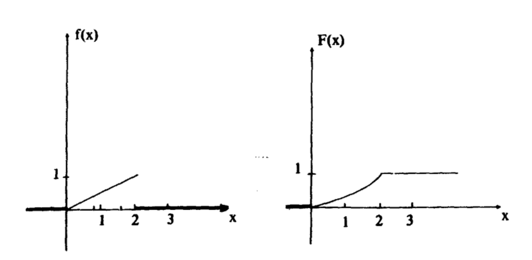
$$F(x) = \int_{-\infty}^{x} f(u)\, du$$

وهنا نميز الحالات الثلاث التالية $2 \le X \le 0$ وأخيرا $x > 2$

ومنه :

$$F(x) = \begin{cases} 0 & : x < 0 \\ \frac{1}{4}x^2 & : 0 \le x \le 2 \\ 1 & : x > 2 \end{cases}$$

كما نجد أن



تمرين (٧)

ألقيت قطعة نقود متماثلة ومتوازنة ثلاث مرات متتالية ، فبفرض أن X يمثل عدد الصور التى ظهرت فى الرمية الأولى ، Y عدد الصور التى ظهرت فى الرميات الثلاث . أوجد توزيع Y, X ، ثم التوزيع المشترك للمتغيرين (X, Y) .

الحل

نلاحظ أن فضاء العينة يتألف من ثمانى نقاط هى :

$$S = (HHH, HTT, THT, TTH, HHT, HTH, THH, TTT)$$

أما قيم المتغير X فهى 0 أو 1 حسبما يظهر كتابة أو صورة على الترتيب ، والاحتمالات الموافقة لهذه القيم فهى $\frac{1}{2}$ ، $\frac{1}{2}$ على التتالى نظراً لتوازن وتماثل قطعة النقود . وهكذا نجد أن جدول توزيع المتغير X هو :

X	0	1
g(x)	$\frac{1}{2}$	$\frac{1}{2}$

كما نلاحظ أن قيم Y الممثل لعدد الصور التى ظهرت فى الرميات الثلاث هى :

$$Y = \{ 0, 1, 2, 3 \} ..$$

أما الاحتمالات الموافقة فتحسب بالعلاقة :

$$h(y) = P [Y = y]$$

والملاحظ أن :

$$h (0) = P [(TTT)] = \frac{1}{8}$$

$$h (1) = P [(HTT, THT, TTH)] = \frac{3}{8}$$

$$h(2) = \frac{3}{8} , h(3) = \frac{1}{8}$$

وبترتيب هذه القيم بجدول نجد أن :

Y	0	1	2	3
h(y)	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$

أما التوزيع الاحتمالى المشترك لـ (X, Y) فهو :

Y \ X	0	1	المجموع
0	$\frac{1}{8}$	0	$\frac{1}{8}$
1	$\frac{2}{8}$	$\frac{1}{8}$	$\frac{3}{8}$
2	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{3}{8}$
3	0	$\frac{1}{8}$	$\frac{1}{8}$
المجموع	$\frac{4}{8}$	$\frac{4}{8}$	1

تمرين (٨)

متغير عشوائى Y دالة كثافته من الشكل :

$$f(y) = \begin{cases} \alpha & : a \le y \le b \\ 0 & : \text{فيما عدا ذلك} \end{cases}$$

(١) حدد α

(٢) أوجد دالة التوزيع Y ((F (y))

(٣) مثل بيانياً كلاً من : $F(y), f(y)$

الحل

من خواص دالة الكثافة نعلم أن :

$$\int_{-\infty}^{+\infty} f(y)\, dy = 1$$

لذلك فإن

$$\int_{0}^{1} \alpha .\, dy = 1$$

ومنه :

$$\alpha = \frac{1}{b-a}$$

وبالتعويض فى عبارة دالة الكثافة نجد أنها تأخذ الشكل التالى :

$$f(y) = \begin{cases} \frac{1}{b-a} & : a \le y \le b \\ 0 & : \text{فيما عدا ذلك} \end{cases}$$

من تعريف دالة التوزيع F نجد أن :

$$F(y) = \int_{-\infty}^{y} f(u)\, du$$

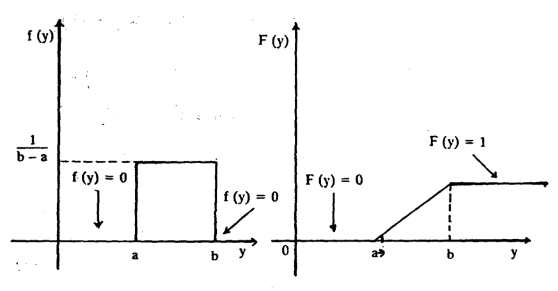
وهنا نميز الحالات الثلاث التالية :

$$y > b,\ a \le y \le b,\ y < a$$

فنجد أن :

$$F(y) = \begin{cases} 0 & : y < a \\ \frac{y-a}{b-a} & : a \le y \le b \\ 1 & : y > b \end{cases}$$

وأخيرا فإننا نجد أن :



تمرين (٩)

بفرض أن الكثافة المشتركة للمتغيرين العشوائيين (X, Y) هى من الشكل :

$$f(x,y) = \begin{cases} 8xy & : 0 < x < 1 \\ & : 0 < y < x \\ 0 & : \text{فيما عدا ذلك} \end{cases}$$

فالمطلوب البحث عن :

$$P\left[Y < \frac{1}{8} \,|\, X = \frac{1}{2} \right], f(y/x)\ ,\ h(y)\ ,\ g(x)$$

الحل

نلاحظ أن :

$$g(x) = \int_{-\infty}^{+\infty} f(x, y)\, dy = \int_{0}^{x} 8x.y\, dy$$

ومنه :

$$g(x) = \begin{cases} 4x^3 & : 0 < x < 1 \\ 0 & : \text{فيما عدا ذلك} \end{cases}$$

كذلك فإن :

$$h(y) = \int_{-\infty}^{+\infty} f(x, y)\, dx = \int_{y}^{1} 8x\, y\, dx$$

ومنه :

$$h(xy = \begin{cases} 4y(1 - y^2) & : 0 < y < 1 \\ 0 & : \text{فيما عدا ذلك} \end{cases}$$

ولكن نعلم أن :

$$f(y \mid x) = \frac{f(x, y)}{g(x)}$$

ولذلك فإن :

$$f(y \mid x) = \frac{2y}{x^2} : 0 < y < x \;,\; 0 < x < 1$$

وكذلك فإن :

$$P[Y < 1/8 \mid X = 1/2] = \int_{0}^{1/8} 8y\, dy = \frac{1}{16}$$

تمرين (١٠)

Y, X متغيران عشوائيان بالتوزيع المشترك :

X / Y	2	3	4	المجموع
1	0.06	0.15	0.09	0.3
2	0.14	0.35	0.21	0.7
المجموع	0.2	0.5	0.3	1

ابحث فيما إذا كان المتغيران السابقان مستقلين أم لا ؟

الحل :

نلاحظ أن جدول توزيع كل من المتغيرين Y, X هو من الشكل :

X	2	3	4
g(x)	0.2	0.5	0.3

X	1	2
g(X)	0.3	0.7

كما نلاحظ أن :

$$g(2)\,.\,h(1) = 0.2 \times 0.3 = 0.06 = f(2.1)$$

$$g(2)\,.\,h(2) = 0.2 \times 0.7 = 0.14 = f(2.2)$$

$$g(3)\,.\,h(1) = 0.5 \times 0.3 = 0.15 = f(3.1)$$

$$g(3)\,.\,h(2) = 0.5 \times 0.7 = 0.35 = f(3.2)$$

$$g(4)\,.\,h(1) = 0.3 \times 0.3 = 0.09 = f(4.1)$$

$$g(4)\,.\,h(2) = 0.3 \times 0.7 = 0.21 = f(4.2)$$

من العلاقات السابقة نجد أنه من أجل أى (x,y) فإن $f(x,y) = g(x)\,.\,h(y)$ مما ينتج معه أن المتغيرين السابقين مستقلان .

تمرين (١١)

بفرض أن التوزيع المشترك للمتغيرين X, Y هو من الشكل :

أوجد توزيع كلا من Y, X والتوزيعات الشرطية ؟

Y \ X	1	2	3	المجموع
1	0	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{3}{12}$
2	$\frac{1}{5}$	$\frac{1}{9}$	0	$\frac{14}{45}$
3	$\frac{2}{15}$	$\frac{1}{4}$	$\frac{1}{18}$	$\frac{79}{180}$
المجموع	$\frac{5}{15}$	$\frac{19}{36}$	$\frac{5}{36}$	1

الحل

نلاحظ أن توزيع المتغير X هو :

X	1	2	3
g(x)	$\frac{1}{3}$	$\frac{19}{36}$	$\frac{5}{36}$

كما أن توزيع Y هو من الشكل :

Y	1	2	3
h(y)	$\frac{1}{4}$	$\frac{14}{45}$	$\frac{79}{180}$

ثم إن التوزيع الشرطي للمتغير X بمعلومية أن Y قد أخذ قيمة يحسب بالعلاقة :

$$f(X \mid Y = y) = \frac{f(x, y)}{h(y)}$$

فمثلا :

$$f(1 \mid 1) = \frac{f(1, 1)}{h(1)} = \frac{0}{1/4} = 0$$

$$f(2 \mid 1) = \frac{f(2, 1)}{h(1)} = \frac{1/6}{1/4} = \frac{2}{3}$$

$$f(3 \mid 1) = \frac{f(3, 1)}{h(1)} = \frac{1/12}{1/4} = \frac{1}{3}$$

$$f(2 \mid 1) = \frac{f(2, 1)}{h(1)} = \frac{1/6}{1/4} = \frac{2}{3}$$

وهكذا بالنسبة لبقية القيم .

وبنفس الطريقة نحسب التوزيع الشرطى للمتغير Y بمعلومية أن X قد أخذ قيمة :

$$f(Y X = x) = \frac{f(x, y)}{g(x)}$$

فمثلا نجد أن :

$$f(1 \mid 3) = \frac{f(3, 1)}{g(3)} = \frac{1/12}{5/36} = \frac{3}{5}$$

$$f(2 \mid 3) = \frac{f(3, 2)}{g(3)} = \frac{0}{5/36} = 0$$

وهكذا ....

تمرين (١٢)

بضاعة مصنعة مؤلفة من 20 قطعة مخلوطة خلطاً جيداً بينها قطعتين معابتين . اخترنا منها أربع قطع بصورة عشوائية ماهى القيمة المتوقعة μ لعدد القطع المعابة ؟

الحل

لنرمز لعدد القطع المعابة والموجودة فى العينة المسحوبة بالرمز X . نلاحظ أن قيم هذا المتغير هى 0, 1, 2 . كما نلاحظ أن فضاء العينة S فى عملية سحب أربع قطع من بين 20 قطعة يتكون من :

$4845 = \binom{20}{4}$ قطعة ، كما نلاحظ أيضا أن الاحتمال الموافق لعدد القطع المعابة المسحوبة X هو :

$$P[X = x] = \frac{\binom{2}{x}\binom{18}{4-x}}{\binom{20}{4}}$$

وهكذا نجد أن القيمة المتوقعة لعدد القطع المعابة هو :

$$\mu = E(X) = \sum_{x=0}^{2} x . P[X = x]$$

$$= \frac{0.\binom{2}{0}\binom{18}{4}}{\binom{20}{4}} + 1\,\frac{\binom{2}{1}\binom{18}{3}}{\binom{20}{4}} + \frac{2\binom{2}{2}\binom{18}{2}}{\binom{20}{4}}$$

ومنه :

$$\mu = 0 + 1.\ \frac{2\,(816)}{4845} + 2.\ \frac{1\,(153)}{4845}$$

$$\mu = 1938/4845 = 0.4$$

تمرين (١٣)

ابحث عن التوقع الرياضى لكل من المتغيرات X. Y, Y. X الواردة فى التمرين (٢،١٠) ثم اثبت أن توقع X. Y يساوى حاصل جُداء توقع X فى توقع Y

الحل

من جدول توزيع X نجد أن :

$$\mu_x = \sum_x x.\ g(x) = 2(0.2) + 3(0.5) + 4(0.3) = 3.1$$

ومن جدول توزيع المتغير Y نجد أيضا أن :

$$\mu_y = \sum_y y.\ h(y) = 1(0.3) + 2(0.7) = 1.7$$

ومن المعلوم أن :

$$\mu_{x.y} = E[X.Y]$$
$$= \sum_x \sum_y x.y.f(x.y)$$

ومن جدول التوزيع Y, X المشترك نجد كذلك أن :

$$\mu_{x.y} = (2.1)\,f(2,1) + (2.2)\,f(2,2) + (3.1)\,f(3,1) + (3.2)\,f(3,2) + (4.1)\,f(4,1) + (4.2)\,f(4,2)$$

وبالتعويض عن f(x,y) بقيمتها من الجدول نجد :

$$\mu_{x.y} = (2.1)(0.06) + (2.2)(0.14) + (3.1)(0,15) + (3.2)(0.35) + (4.1)(0.09) + (4.2)(0.21)$$
$$= 0.12 + 0.56 + 0.45 + 2.1 + 0.36 + 1.68$$
$$\mu_{x.y} = 5.27$$

نلاحظ مما سبق أن :

$$\mu_x \,.\, \mu_y = (3.1)\,.\,(1.7) = 5.27$$

وبمقارنة النتيجتين الأخيرتين نجد المطلوب .

تمرين (١٤)

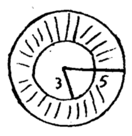
يرمى رامٍ على هدف مكون من دائرتين متمركزتين نصفى قطريهما على الترتيب 3, 5 فإذا أصاب الرامى الدائرة الداخلية فإنه يسجل له العدد 10 ، أما إذا أصاب القسم المظلل فإنه يسجل له 5 . فإذا ، علمت أن احتمال إصابته للهدف هو 0.70 ، وأن احتمالات إصابته لأى نقطة من الهدف متساوية . فما هى القيمة المتوقعة μ لعدد النقاط التى يسجلها فى كل رمية ؟

الحل

من الملاحظ أن احتمال حصول الرامى على 10 أو 5 من النقاط على الترتيب هو :

$$f(10) = 0.7 \frac{\text{مساحة الدائرة الصغرى}}{\text{مساحة الهدف}} = 0.7 \frac{3^2}{5^2} = 0.252$$

$$f(5) = 0.7 \frac{\text{مساحة الحلقة المظللة}}{\text{مساحة الهدف}} = 0.7 \frac{(5^2 - 3^2)}{5^2} = 0.448$$



ونلاحظ أن القيمة المتوقعة لعدد النقاط التى يسجلها فى كل رمية هى :

$$\mu = 5.(0.448) + 10.(0.252) = 4.76$$

تمرين (١٥)

ما هو التوقع الرياضى للعدد الملاحظ على حجر نرد عند إلقائه ؟

الحل

إذا رمزنا للعدد الملاحظ بالرمز X ، فإننا نجد أن جدول توزيع X هو :

X	1	2	3	4	5	6
f(x)	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

ومنه :

$$\mu = E(X) = \frac{1}{6}.1 + \frac{1}{6}.2 + \frac{1}{6}.3 + \frac{1}{6}.4 + \frac{1}{6}.5 + \frac{1}{6}.6$$

$$= 3.5$$

تمرين (١٦)

بفرض أن X يمثل الوجه الذى ظهر عند إلقاء حجر نرد متوازن ومتماثل . أوجد توقع المتغير $Y = 2X^2 - 5$ ؟

الحل

نعلم أن جدول توزيع المتغير X هو من الشكل :

X	1	2	3	4	5	6
f(x)	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

كما أن :

X^2	1	4	9	16	25	36
$f(x^2)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

والملاحظ أن جدول توزيع المتغير Y هو من الشكل :

$Y = 2x^2$	-3	3	13	27	45	67
$f(2x^2)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

وهكذا نجد أن :

$$E[Y] = \frac{1}{6}[-3 + 3 + 13 + 27 + 45 + 67] = \frac{1}{6}(152)$$

وأخيراً فإن :

$$E[Y] = \frac{76}{3}$$

تمرين (١٧)

فى المثال (٢,٢١) احسب التوقع الرياضى للمتغير X^2-1 ، ثم احسب تباين المتغير X .

الحل

نعلم أن جدول توزيع المتغير X هو من الشكل :

X	-1	0	1	2
f(x)	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{3}{8}$	$\frac{1}{4}$

نلاحظ أن :

$$E(X^2-1) = \sum_x (x^2-1)\, f(x)$$

$$= 0.\frac{1}{8} + (-1).\frac{1}{4} + 0.\frac{3}{8} + 3.\frac{1}{4} = \frac{1}{2}$$

ومن خواص التوقع الرياضى نجد أن :

$$E(X^2) - 1 = \frac{1}{2}$$

ومنه :

$$E(X^2) = \frac{3}{2}$$

وأخيراً فإن تباين المتغير X يعطى بالعلاقة :

$$\sigma^2 = E(X^2) - \mu^2 = \frac{3}{2} - \left(\frac{3}{4}\right)^2$$

$$\sigma^2 = 1.3125$$

هنا عوضنا عن $\mu = \frac{3}{4}$ كما وجدنا فى المثال (٢,٢١)

تمرين (١٨)

جهاز إرسال يحوى سبع ترانزستورات ، اثنان منها عاطلان . اخترنا بصورة عشوائية ثلاثة ترانزستورات ، ونزعناها من جهاز الإرسال ، وفحصناها . ما هو التوقع الرياضى لعدد الترانزستورات العاطلة المنزوعة ؟

الحل

لنرمز بـ Y لعدد الترانزستورات العاطلة المنزوعة .

نلاحظ أن جدول قانون توزيع Y هو من الشكل :

X	0	1	2
f(x)	$\frac{10}{35}$	$\frac{20}{35}$	$\frac{5}{35}$

ذلك لأن :

$$P[Y = y] = \frac{\binom{2}{y}\binom{5}{3-y}}{\binom{7}{3}}$$

وهكذا نجد أن توقع Y هو :

$$\mu_y = E(Y) = 0.\left(\frac{10}{35}\right) + 1.\left(\frac{20}{35}\right) + 2.\left(\frac{5}{35}\right) = 0.857$$

ولحساب التباين نعلم أن :

$$\sigma_y^2 = E(Y^2) - \mu_y^2$$

غير أن :

$$E(Y^2) = \sum_y y^2 . h(y)$$

$$= 0.\left(\frac{10}{35}\right) + 1.\left(\frac{20}{35}\right) + 4.\left(\frac{5}{35}\right)$$

$$= \frac{40}{35} = 1.1428$$

بالتعويض نجد أن :

$$\sigma_y^2 = 1.1428 - (0.857)^2$$

$$= 0.408$$

تمرين (١٩)

X متغير عشوائى مستمر بالكثافة الاحتمالية :

$$f(x) = \begin{cases} k\,e^{-x} & : \quad x \geq 0 \\ 0 & : \quad \text{فيما عدا ذلك} \end{cases}$$

والمطلوب :

١ ـــ البحث عن قيمة الثابت K ؟

٢ ـــ توقع المتغير X ؟

٣ ـــ تباين هذا المتغير ؟

الحل

نلاحظ من خواص الكثافة الاحتمالية أن :

$$\int_{-\infty}^{+\infty} f(x)\,dx = 1 \Leftrightarrow \int_{0}^{\infty} K e^{-x}\,dx = 1 = K = 1$$

والكثافة الاحتمالية بعد تعويض K = 1 هى من الشكل :

$$f(x) = \begin{cases} e^{-x} & : \quad x \geq 0 \\ 0 & : \quad \text{فيما عدا ذلك} \end{cases}$$

أما التوقع الرياضى فى هذه الحالة فيحسب بالعلاقة :

$$\mu = E(X) = \int_{-\infty}^{+\infty} x.\, f(x)\, dx = \int_{0}^{\infty} x.\, e^{-x}\, dx$$

وبالمكاملة بالتجزئة باعتبار أن $u = x, dv = e^{-x}\, dx$ نجد أن :

$$\mu = -x\, .\, e^{-x} \Big)_{x=0}^{\infty} + \int_{0}^{\infty} e^{-x}\, .\, dx$$

$$\mu = 0 + 1 = 1$$

ولحساب التباين نبدأ بحساب $\mu_2 = E(X^2)$ فنجد أن :

$$\grave{\mu}_2 = E(X^2) = \int_{0}^{\infty} x^2\, .\, e^{-x}\, dx$$

وبالمكاملة بالتجزئة على أساس أن $u = x^2, dv = e^{-x}\, dx$ نجد أن :

$$\grave{\mu}_2 = x^2\, .\, e^{-x} \Big)_{x=0}^{\infty} + 2 \int_{0}^{\infty} x\, e^{-x}\, dx$$

ومن الملاحظ أن $1 = \int_{0}^{\infty} x\, e^{-x}\, dx = \mu$ وهكذا نجد :

$$\grave{\mu}_2 = E(X^2) = 0 + 2.1 = 2$$

وأخيراً فإن :

$$\sigma^2 = \grave{\mu}_2 - \mu^2 = 2 - 1 = 1$$

تمرين (٢٠)

متغير عشوائى Y توقعه الرياضى $\mu = 10$ ، وانحرافه المعيارى $\sigma = 2$ احسب باستخدام متباينة تشيبشيف :

١ — $P[\,|Y - 10| \geq 3\,]$

٢ — $P[\,|Y - 10| < 3\,]$

٣ — $P[\,5 < Y < 15\,]$

الحل

أولا ــ نلاحظ أن :

$$P[\,|Y-10|\geq 3\,] = 1-P[X-10\,|<3]$$
$$= 1-P[\,10-3<X<10+3\,]$$
$$= 1-P[\,10-\frac{3}{2}\,.\,2< X<10+\frac{3}{2}\,.\,2\,]$$

ولكن نعلم أن :

$$P[\mu-k\cdot\sigma<X<\mu+k\cdot\sigma]\geq 1-\frac{1}{k^2}$$

وبملاحظة أن $k=\frac{3}{2}$ فإننا نجد أن :

$$P[\,|Y-10|\geq 3\,]\leq 1-1+\frac{1}{k^2}$$
$$\leq\frac{4}{9}$$

ثانيا ــ نعلم أن :

$$P[\,|Y-10|<3\,] = 1-P[\,|Y-10|\geq 3\,]$$

وباسخدام نتيجة الطلب الأول نجد أن :

$$P[\,|Y-10|\geq 3\,]\geq\frac{5}{9}$$

ثالثا ــ نلاحظ أن :

$$P[\,5<Y<15\,] = P[\,10-5<Y<10+5\,]\geq 1-\frac{1}{k^2}$$

حيث يمثل $5 = 2K = K\,.\,\sigma$ ومنه $K=\frac{5}{2}$

وأخيراً فإن :

$$P[5<Y<15]\geq\frac{21}{25}$$

تمرين (٢١)

احسب $P[\mu - 2\sigma < X < \mu + 2\sigma]$ إذا علمت أن للمتغير X توزيعاً معطى بالكثافة :

$$f(x) = \begin{cases} 6x(1-x) & : 0 < x < 1 \\ 0 & : \text{فيما عدا ذلك} \end{cases}$$

الحل

نبدأ بحساب μ, $E(X^2)$ ثم σ^2 فنجد أن :

$$\mu = \int_0^1 x.[6x(1-x)]\,dx = \frac{1}{2}$$

$$, E(X^2) = \int_0^1 x^2[6x(1-x)]\,dx$$

$$E(X^2) = \frac{3}{10}$$

وهكذا نجد أن :

$$\sigma^2 = E(X^2) - \mu^2 = \frac{3}{10} - \frac{1}{4} = \frac{1}{20}$$

وأخيرا فإن :

$$2\sigma = \frac{1}{\sqrt{5}} = 0.4472$$

ومنه :

$$P[\mu - 2\sigma < X < \mu + 2\sigma] = P[0.5 - 0.447 < X < 0.5 + 0.447]$$

$$P[0.053 < X < 0.947]$$

$$\int_{0.053}^{0.947} 6x\,(1 - x)\,dx$$

$$= 0.98$$

تمرين (٢٢)

متغيران عشوائيان منقطعان جدول توزيعهما المشترك من الشكل :

Y \ X	-2	-1	0	1	2	3	
0	0.05	0.05	0.1	0	0.05	0.05	0.30
1	0.10	0.05	0.05	0.1	0	0.05	0.35
2	0.03	0.12	0.07	0.06	0.03	0.04	0.35
	0.18	0.22	0.22	0.16	0.08	0.14	1

والمطلوب :

١ ـــ ايجاد توزيع $Z = X + Y$

٢ ـــ حساب التوقع الرياضى لـ Z والتحقق من أن $E(Z) = E(X) + E(Y)$

الحل

من الواضح أنه إذا رمزنا بـ ϕ للتوزيع الاحتمالى للمتغير Z لوجدنا أن قيم هذا المتغير هى :

$$Z: -2, -1, 0, 1, 2, 3, 4, 5$$

كما أن قيم التوزيع الاحتمالى فى هذه النقاط هى :

$\phi(-2) = 0.05$

$\phi(-1) = 0.05 + 0.1 = 0.15$

$\phi(0) = 0.1 + 0.05 + 0.03 = 0.18$

$\phi(1) = 0 + 0.05 + 0.12 = 0.17$

$\phi(2) = 0.05 + 0.1 + 0.07 = 0.22$

$\phi(3) = 0.05 + 0 + 0.06 = 0.11$

$\phi(4) = 0.05 + 0.03 = 0.08$

$\phi(5) = 0.04$

وهكذا نستنتج أن جدول توزيع المتغير الجديد Z هو من الشكل :

Z	-2	-1	0	1	2	3	4	5
$\phi(Z)$	0.05	0.15	0.18	0.17	0.22	0.11	0.08	0.04

أما التوقع الرياضى للمتغير Z فيحسب بالشكل التالى :

$$\mu_z = E(Z) = (-2)(0.05) + (-1)(0.15) + (0)(0.18) + (1)(0.17) + (2)(0.22) + (3)(0.11) + (4)(0.08) + (5)(0.04)$$

ومنه :

$$\mu_z = 1.21$$

غير أن :

$$\mu_x = (-2)(0.18) + (-1)(0.22) + (0)(0.22) + (1)(0.16) + (2)(0.08) + (3)(0.14)$$

ومنه :

$$\mu_x = 0.16$$

كذلك فإن :

$$\mu_Y = (0)(0.3) + (1)(0.35) + (2)(0.35) = 1.05$$

وأخيراً فإننا نلاحظ أن :

$$\mu_z = \mu_X + \mu_Y$$

تمرين (٢٣)

يجرى الرمى على هدف بوساطة سلاح معين . فإذا علمت أن احتمال إصابة الهدف P ، وأن X, Y يمثلان على الترتيب عدد مرات أخطاء ، واصابة الهدف ، فالمطلوب إيجاد دالة التوزيع المشتركة لهذين المتغيرين ؟

الحل

نلاحظ أن المتغيرين العشوائيين X, Y مرتبطان بالعلاقة التابعية :

$$X + Y = 1$$

لذلك فإن التوزيع المشترك لهذين المتغيرين يوضحه الجدول التالى :

Y \ X	0	1
0	0	q
1	p	0

تمرين (٢٤)

يصوب راميان على هدف ل كل بسلاحه وبصورة مستقلة عن الآخر ، فإذا علمت أن عدد المرات التى أصاب بها الرامى الأول الهدف هو X ، والرامى الثانى هو Y ، وأن P_1 احتمال اصابة الهدف من قبل الرامى الأول و P_2 من قبل الرامى الثانى ، فأوجد دالة التوزيع المشترك للمتغيرين X, Y ؟

الحل

بما أن كل رام يصوب على الهدف بصورة مستقلة عن الآخر لذلك فإن :

$$F(x, y) = P(X < x).P(Y < Y)$$
$$= F_1(x).F_2(y)$$

ولكن :

$$F_1(x) = \begin{cases} 0 & : x \leq 0 \\ q_1 = 1 - P_1 & : 0 < x \leq 1 \\ 1 & : x > 1 \end{cases}$$

وكذلك فإن :

$$F_2(y) = \begin{cases} 0 & : y \leq 0 \\ q_2 = 1 - P_2 & : 0 < y \leq 1 \\ 1 & : y > 1 \end{cases}$$

لذلك فإن جدول التوزيع المشترك لهذين المتغيرين هو :

Y \ X	$x \leq 0$	$0 < x \leq 1$	$x > 1$
$y \leq 0$	0	0	0
$0 < y \leq 1$	0	$q_1 q_2$	q_2
$1 < y$	0	q_1	1

تمرين (٢٥)

يجرى التصويب على هدف بسلاحين مستقلين واحتمال اصابة الهدف بكل سلاح هو P ، فإذا علمت أن X يمثل الفرق بين عدد الاصابات وعدد الطلقات التى أخطأت الهدف ، وأن Y يمثل مجموع عدد الاصابات وعدد الطلقات التى أخطأت الهدف فالمطلوب :

(١) البحث عن توزيع كل من Y,X

(٢) حساب $E(X), E(Y), D(X), D(Y)$ ؟

الحل

نلاحظ أن :

X	-2	0	2
P(x)	q^2	2pq	p^2

Y	2
P(Y)	1

لذلك وحسب تعريف التوقع الرياضى ، والتباين نجد أن :

$$E(X) = 2(p^2 - q^2) = 2(p - q)$$

$$E(Y) = 2(1) = 2$$

$$D(X) = 4(p^2 + q^2) - 4(p^2 - 2pq + q^2) = 8pq$$

$$D(Y) = 0$$

تمرين (٢٦)

يرمى راميان على هدف معين كل بسلاحه وبصورة مستقلة عن الآخر . فإذا علمت أن احتمال إصابة الأول للهدف هو P_1 ، والثانى هو P_2 وأن X يمثل عدد المرات التى أصاب بها الأول الهدف ، Y عدد المرات التى أصاب بها الثانى الهدف ، وأن Z يمثل الفرق بين عدد مرات الإصابة . فالمطلوب هو إيجاد قانون توزيع Z وتوقعه الرياضى

الحل

نلاحظ أن للمتغير Z ثلاث قيم هى 1 + ,1,0 - ، وأن :

$P(Z = -1) = P(X = 0)\, P(Y = +1) = q_1p_2$

$P(Z = 0) = P(X = 0)\, P(Y = 0) + P(X = 1)\, P(Y = 1) = q_1q_2 + p_1p_2$

$P(Z = 1) = P(X = 1)\, P(Y = 0) = P_1q_2$

Z	−1	0	1
P(z)	q_1p_2	$q_1q_2 + p_1p_2$	P_1q_2

من الجدول السابق نستنتج أن :

$$E(Z) = -q_1p_2 + p_1q_2 = p_1 - p_2$$

تمرين (٢٧)

بفرض أن دالة الكثافة المشتركة للمتغيرين المستمرين X, Y هى من الشكل :

$$f(x, y) = \begin{cases} \frac{1}{\pi r^2} & : x^2 + y^2 < r \\ 0 & : x^2 + y^2 > r \end{cases}$$

فأوجد دالة الكثافة الشرطية للمتغير X ، علما بأن المتغير Y قد أخذ قيمة y وأن $0 < |y| < r$

الحل

من المعلوم أن :

$$f(x|Y = y) = \frac{f(x, y)}{f(y)}$$

ومن الفرض لدينا :

$$= \frac{\frac{1}{\pi r^2}}{\frac{1}{\pi r^2} \int_{-\sqrt{r^2 - y^2}}^{\sqrt{r^2 - y^2}} dx}$$

وبما أن :

$$= \frac{1}{2\sqrt{r^2 - y^2}} \quad : \quad |x| < \sqrt{r^2 - y^2}$$

$$f(x, y) = 0 \text{ for } x^2 + y^2 > r^2$$

لذلك فإن :

$$f(x \mid Y = y) = 0 \quad \text{for} \quad |x| > \sqrt{r^2 - y^2}$$

تمرين (٢٨)

بفرض أن دالة الكثافة المشتركة للثنائية (X, Y) هى من الشكل :

$$f(x, y) = \begin{cases} \frac{1}{6\pi} & : \frac{x^2}{9} + \frac{y^2}{4} < 1 \\ 0 & : \frac{x^2}{9} + \frac{y^2}{4} > 1 \end{cases}$$

فابحث عن دالة الكثافة لكل من X, Y ؟

الحل :

$$f(x) = \int^{+\infty} f(x, y)dy = \frac{1}{6\pi} \int_{-2\sqrt{1 - \frac{x^2}{9}}}^{2\sqrt{1 - \frac{x^2}{9}}} dy$$

ومنه

$$g(x) = \begin{cases} \frac{2}{9\pi} \sqrt{9 - x^2} & : |x| < 3 \\ 0 & : |x| \geq 3 \end{cases}$$

بنفس الطريقة نجد أن :

$$h(y) = \begin{cases} \frac{1}{2\pi} \sqrt{4 - y^2} & : |y| < 2 \\ 0 & : |y| \geq 2 \end{cases}$$

تمرين (٢٩)

بفرض أن الجدول الثاني يمثل التوزيع المشترك للمتغيرين X. Y :

Y \ X	x_1	x_2	x_3
y_1	0.1	0.3	0.2
y_2	0.6	0.18	0.16

أوجد دالة الكثافة الشرطية للمتغير X إذا علمت أن المتغير Y قد أخذ القيمة y_1 .

الحل

نعلم أن :

$$f(x_1 \mid Y = y_1) = \frac{P\,[X = x_1 \,,\, Y = y_1]}{h(y_1)}$$

وبالتعويض نجد أن :

$$f(x_2 \mid Y = y_1) = \frac{0.1}{0.6} = \frac{1}{6}$$

كذلك فإن :

$$f(x_2 \mid Y = y_1) = \frac{0.3}{0.6} = \frac{1}{2}$$

$$f(x_3 \mid Y = y_1) = \frac{0.2}{0.6} = \frac{1}{3}$$

تمارين عامة

(١) حدد فيما اذا كانت المتغيرات التالية منقطعة أم مستمرة :

أ – X – عدد حوادث السيارات التى وقعت خلال عام فى المملكة العربية السعودية .

ب – Y – عدد المبانى التى تم إنشاؤها فى مدينة جدة .

جـ – Z – عدد البيضات التى تبيضها دجاجة خلال شهر أيلول .

د – M – قيمة الحليب الذى تنتجه بقرة فى مزارع فقيه للدواجن خلال عام .

(٢) يحتوى صندوق على أربع كرات سود وكرتين خضر . سحبنا ثلاث كرات من الصندوق على التتالى ، ومع إعادة الكرة المسحوبة إلى الصندوق قبل سحب الأخرى ، ما هو التوزيع الاحتمالى لعدد الكرات الخضر المسحوبة من هذا الصندوق ؟

(٣) أوجد عبارة التوزيع الاحتمالى للمتغير العشوائى X الممثل للنتيجة التى تظهر عند إلقاء حجر نرد مرة واحدة .

(٤) بضاعة مصنعة مؤلفة من ست قطع من بينها قطعتان معابتان ، أوجد التوزيع الاحتمالى للمتغير X الممثل لعدد القطع المعابة لدى سحب ثلاث قطع من هذه البضاعة ، وذلك بصورة عشوائية . عبر عن النتائج بوساطة ما يسمى بمخطط توزيع التواتر .

(٥) أوجد فى المثال ٤ التوزيع التراكمي للمتغير X . استخدام F(x) للبحث عن :

أ – $P[X = 1]$

ب – $P[0 < X < 2]$

(٦) إذا علمت أن الكثافة الاحتمالية لمتغير عشوائى مستمر X هى من الشكل :

$$f(x) = \begin{cases} k\sqrt{x} & : 0 < x < 1 \\ 0 & \text{: فيما عدا ذلك} \end{cases}$$

فالمطلوب :

أ ــ تحديد قيمة الثابت K

ب ــ إيجاد F(x) ، ثم استخدمه لحساب قيمة $P(0.3 < X < 0.6)$.

(٧) يحتوى صندوق فاكهة على ثلاث برتقالات وتفاحتين وثلاث موزات . سحبنا وبصورة عشوائية أربع قطع من الصندوق . فإذا كان X ممثلا لعدد البرتقالات المسحوبة ، Y ممثلا لعدد التفاحات المسحوبة أيضا فالمطلوب :

أ ــ إيجاد التوزيع الاحتمالي المشترك للمتغيرين (X, Y) .

ب ــ $P[(x, y) \in A]$ حيث تمثل A المنطقة من المستوى المحددة بالمجموعة $\{(x, y) | x + y \leq 2\}$

(٨) متغيران عشوائيان مستمران لهما كثافة احتمالية مشتركة محددة بالعلاقة :

$$f(x,y) = \begin{cases} 4x.y & : 0 < x < 1 , 0 < y < 1 \\ 0 & \text{: فيما عدا ذلك} \end{cases}$$

والمطلوب :

أ ــ ما هو $P[0 \leq x < \frac{3}{4} | \frac{1}{8} \leq y \leq \frac{1}{2}]$

ب ــ ما هو $P[Y > X]$

(٩) متغيران عشوائيان مستمران لهما كثافة احتمالية مشتركة من الشكل :

$$f(x,y) = \begin{cases} K(x^2 + y^2) & : 0 < x < 2 , 1 < y < 4 \\ 0 & \text{: فيما عدا ذلك} \end{cases}$$

فالمطلوب :

أ ــ تحديد قيمة K

ب ــ إيجاد قيمة الاحتمال $P[1 < X < 2 , 2 < Y \leq 3]$

جـ ــ ما هي قيمة $P[1 \leq X \leq 2]$

د ــ إيجاد قيمة الاحتمال $P[X + Y > 4]$

(١٠) متغيران عشوائيان بالكثافة المشتركة :

$$f(x,y) = \begin{cases} \frac{1}{y} & : 0 < x < y , 0 < y < 1 \\ 0 & : \text{فيما عدا ذلك} \end{cases}$$

أوجد : $P[X + Y > \frac{1}{2}]$

(١١) عد إلى التمرين (٥) واحسب :

أ ــ $f(y / 2)$

ب ــ $P[Y = 0|X = 2]$

(١٢) إذا علمت أن التوزيع الاحتمالي المشترك للمتغيرين العشوائيين X, Y معطى بالجدول التالي :

Y \ X	1	2	3
1	0	$\frac{1}{6}$	$\frac{1}{12}$
2	$\frac{1}{5}$	$\frac{1}{9}$	0
3	$\frac{2}{15}$	$\frac{1}{4}$	$\frac{1}{15}$

فأوجد التوزيعات الهامشية والشرطية لهذين المتغيرين .

(١٣) بفرض أن الكثافة المشتركة للمتغيرين المستمرين (X, Y) هي من الشكل :

$$f(x, y) = \begin{cases} \dfrac{6-x-y}{8} & : 0 < x < 2, \ : 2 < y < 4 \\ 0 & \text{فيما عدا ذلك :} \end{cases}$$

أوجد $P[1 < Y < 3 | X = 2]$

(١٤) بفرض أن الكثافة الاحتمالية المشتركة للمتغيرين (X, Y) هى من الشكل :

$$f(x,y) = \begin{cases} 2 & : 0 < x < y < 1 \\ 0 & \text{فيما عدا ذلك :} \end{cases}$$

فالمطلوب :

أ ـــ تحديد فيما إذا كان المتغيران مستقلين .

ب ـــ إيجاد :

$$P\left[\frac{1}{4} < X < \frac{1}{2} \middle| Y = \frac{3}{4}\right]$$

(١٥) متغير عشوائى توزيعه الاحتمالى معطى بالجدول التالى :

X	-3	6	9
P [X = x]	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{1}{3}$

والمطلوب :

أ ـــ حساب قيمة $E(X^2), E(X)$

ب ـــ استخدام قوانين التوقع الرياضى فى حساب :

$$E[(2X + 1)^2], E[\{X - E(X)^2\}]$$

(١٦) بفرض أن X يمثل العدد الذى ظهر لدى إلقاء حجر نرد أحمر ، Y العدد الذى ظهر لدى إلقاء نرد أسود . أوجد :

أ — $E(X + Y)$

ب — $E(X - Y)$

جـ — $E(X . Y)$

(١٧) أوجد تمام تباين المتغيرين العشوائيين فى التمرين (٧) .

(١٨) أوجد تمام المتغيرين العشوائيين فى التمرين (١٢) .

(١٩) برهن أن : $Cov(ax, by) = a.b\, Cov(x, y)$

(٢٠) متغير عشوائى X له متوسط $\mu = 12$ وتباين $\sigma^2 = 9$ ، وتوزيعه الاحتمالى غير معلوم . مستخدماً متباينة تشيبشيف أوجد :

أ — $P[6 < X < 18]$

ب — $P[3 < X < 21]$

(٢١) برهن متباينة تشيبشيف من أجل أى متغير عشوائى X منقطع .

(٢٢) متغير عشوائى مستمر دالة كثافته من الشكل :

$$f(x) = \begin{cases} \frac{1}{2}\cos x & : x \in \left(\frac{-\pi}{2}, \frac{\pi}{2}\right) \\ 0 & : x \notin \left(\frac{-\pi}{2}, \frac{\pi}{2}\right) \end{cases}$$

احسب توقع وتباين هذا المتغير .

الفصل الثالث

بعض توزيعات الاحتمال المنقطعة

■ مقدمة ■ التوزيع المنتظم ■ التوزيع الحدانى والتوزيع المتعـدد الحـدود ■ التوزيع الهندسى الزائدى ■ التوزيع البواسونى ■ التوزيع الحدانى السالب ■ تمارين محلولة ■ تمارين عامة .

(٣,١) مقدمة

لقد أوضحنا فى الفصل الثانى أن توزيع الاحتمال المنقطع يمثل بجدول يوضح قيم المتغير العشوائى ، والاحتمالات الموافقة لهذه القيم . والمسألة المطروحة الآن ما هى الطريقة التى تصف لنا سلوك متغير عشوائى ما . الملاحظ أن بعض المتغيرات العشوائية والتى تتعلق ببعض التجارب الإحصائية تتصف بخواص متشابهة ، ويمكن أن تصف بنفس توزيع الاحتمال ، فمثلا أن جميع المتغيرات العشوائية الممثلة لعدد مرات النجاح فى n اختبارا معادا لتجارب مستقلة (حيث يمثل احتمال الحصول على نجاح فى كل اختبار مقدارا ثابتا لا يتغير من اختبار لآخر) يتصف سلوكها بنفس النموذج ، ولذلك يمكن تمثيلها بصيغة خاصة .

فى هذا الفصل سندرس عددا هاما من توزيعات الاحتمال المنقطعة . إن هذه التوزيعات تصف لنا معظم المتغيرات العشوائية التى تصادفنا خلال تجاربنا العملية .

(٣,٢) التوزيع المنتظم Uniform distribution

إن من أبسط توزيعات الاحتمال المنقطعة هو ذاك التوزيع الذى يفترض فيه المتغير العشوائى جميع قيمه باحتمالات متساوية . مثل هذا التوزيع يدعى بالتوزيع المنتظم .

التوزيع المنتظم

إذا افترض المتغير العشوائى X جميع القيم $x_1 , x_2 , \dots , x_n$ باحتمالات متساوية ، فإننا نعرف عندئذ توزيع الاحتمال لهذا المتغير المنتظم بالعلاقة :

$$f(x ; n) = \frac{1}{n} : x = x_1 , x_2 , \dots , x_n$$

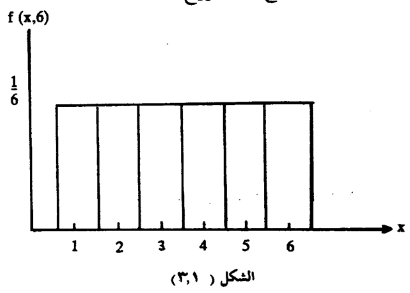
لقد استخدمنا الرمز $f(x ; n)$ بدلا عن $f(x)$ لنشير إلى أن التوزيع المنتظم يتعلق بـ n (عدد القيم التى يفترضها المتغير العشوائى) .

مثال (٣,١)

عند إلقاء حجر نرد متوازن ومتماثل فإننا نجد أن كل نتيجة فى فضاء العينة $S = \{1, 2, ..., 6\}$ يمكن أن تظهر باحتمال قدره $\frac{1}{6}$ لذلك فإن المتغير العشوائى الممثل للعدد الذى سيظهر يمثل متغيراً عشوائياً منتظماً .

الحل

إن المنحنى البيانى الممثل لتوزيع المنتظم يمثل مجموعة مستطيلات بارتفاعات متساوية . والشكل (٣,١) يوضح هذا التوزع .



الشكل (٣,١)

نظرية (٣,١)

توقع وتباين متغير عشوائى منتظم توزيعه $f (x ; n)$ يعطى بالعلاقتين التاليتين :

$$\mu = \frac{\sum_{i=1}^{n} x_i}{n} \quad , \quad \sigma^2 = \frac{\sum_{i=1}^{n} (x_i - \mu)^2}{n}$$

البرهان

من تعريف التوقع الرياضى لمتغير عشوائى نكتب :

$$\mu = E(X) = \sum_{i=1}^{n} x_i \cdot f(x_i\,;\,n) = \sum_{i=1}^{n} \frac{x_i}{n}$$

$$= \frac{\sum^{n} x_i}{n}$$

وحسب تعريف التباين نجد أن :

$$\sigma^2 = E[(X-\mu)^2] = \sum_{i=1}^{n} (x_i-\mu)^2 f(x_i\,;\,n)$$

$$= \sum_{i=1}^{n} \frac{(x_i-\mu)^2}{n} = \frac{\sum^{n} (x_i-\mu)^2}{n}$$

مثال (٣,٢)

احسب فى المثال (٣,١) توقع وتباين المتغير الممثل للوجه الذى ظهر لدى إلقاء قطعة زهر .

الحل

حسب النظرية (٣,١) لدينا :

$$\mu = \frac{\sum^{n} x_i}{n} = \frac{1+2+3+4+5+6}{6} = 3.5$$

$$\sigma^2 = \frac{(1-3.5)^2 + (2-3.5)^2 + \ldots + (6-3.5)^2}{6} = \frac{35}{12}$$

(٣,٣) التوزيع الحدّانى والتوزيع المتعدد الحدود

Binomial and multinomial distribution

يقترن أحد أهم التوزيعات العشوائية المنقطعة بتجربة إلقاء قطعة النقود التى ذكرناها فى الأمثلة (١,٦) ، (١,١٨) ، (١,٢١) ، (١,٢٦)،... ،إلخ وتتكرر يوميا العديد من التجارب التى تشبه تجربة إلقاء النقود ذات الأهمية التطبيقية فى العلوم المختلفة .

فالرمى على هدف يشبه إلى حد كبير إلقاء قطعة النقود لأن عملية الرمى تقود إلى إحدى نتيجتين إما إصابة للهدف أو إخطاء له ، كذلك الأمر فى فحص فعالية دواء جديد فإما أن يكون هذا الدواء فعالا أو غير فعال ، كذلك عند معرفة رأى ناخب فى مرشح ما ، إما أن يصوت ضد أو مع هذا المرشح ، وأخيراً فإن فحص قطعة من بضاعة مصنعة سيترك لنا الباب مفتوحا للحكم على نوع هذه القطعة إن كانت معابة أو جيدة الصنع . كل هذه الأمثلة وغيرها تكشف لنا أن هذه التجارب تتشابه إلى حد مقبول فى الخواص التالية :

(١) تتألف كل تجربة من عدد n من الاختبارات المتماثلة تماماً .

(٢) كل تكرار للتجربة ينتج عنه إحدى نتيجتين إما نجاح أو فشل . فسؤالنا عن فعالية دواء مثلا ضد مرض معين سيتحدد بإحدى النتيجتين ، فعال (نجاح) أو غير فعال (فشل) كذلك إصابة رامٍ للهدف يمثل نجاحاً ، وإخطاؤه للهدف فشلا .

(٣) إن احتمال النجاح فى كل اختبار يبقى ثابتا. فمثلاً يصيب رامى الهدف باحتمال قدره 0.7 ، وهذا الاحتمال لا يتغير سواء فى الرمية الأولى أو العاشرة . كذلك فإن احتمال مقابلة ناخب مؤيد للمرشح الفلانى ، يبقى ثابتا تقريبا طالما كان مجتمع الناخبين كبيرا جدا بالمقارنة مع العينة من الناخبين الذين تجرى مقابلتهم . فإذا كان 50% مثلا من المجتمع الأمريكى يحوى ألف ناخب يفضلون المرشح إدوارد كنيدى ، فإن احتمال الحصول على تأييد لهذا المرشح عند أول مقابلة هو $\frac{1}{2}$. واحتمال التأييد عند المقابلة الثانية هو $\frac{500}{999}$ أو $\frac{499}{999}$ حسبما تكون المقابلة الأولى قد تمت مع مؤيد أو مع معارض على الترتيب . والعددان قريبان جدا من $\frac{1}{2}$ ويمكن اعتبارهما $\frac{1}{2}$ عمليا ، كما يمكن أن نستمر فى مثل هذا والرابعة حتى المقابلة الـ n طالما ظلت n صغيرة بالنسبة للعدد 1000 ، وبصورة خاصة إذا كان عدد الناخبين عشرة ، وكان خمسة منهم يفضلون المرشح إدوارد كنيدى ، فإن احتمال الحصول على تأييد فى المقابلة الأولى هو $\frac{1}{2}$ ، أما فى الثانية فهو $\frac{4}{9}$ أو $\frac{5}{9}$. أى أن الاحتمال P يتغير كثيراً من اختبار إلى آخر ، والتجربة لن تكون تجربة حدانية .

(٤) الاختبارات المكررة مستقلة .

وبذلك تتمتع التجربة الحدانية بالخواص الأربعة التالية :

(١) تتألف كل تجربة من n اختبار معاداً .

(٢) إن نتيجة كل اختبار إما نجاح أو فشل .
(٣) احتمال النجاح فى كل اختبار ثابت لا يتغير من اختبار لآخر .
(٤) الاختبارات المكررة مستقلة .

لنفرض على سبيل المثال أن تجربتنا تنحصر فى سحب عنصرين من مجموعة بضاعة مصنعة بشكل عشوائي . سنرمز بـ N للعنصر المسحوب من نوع جيد و بـ D للعنصر المعاب ، وسنفرض أن العنصر المعاب هو بمثابة نجاح . أن عدد النجاحات X هو متغير عشوائى يفترض القيم 0, 1, 2, وفى هذه التجربة نجد النتائج الأربع التالية :

النتيجة	X
NN	0
DN	1
ND	1
DD	2

لنفرض أن عدد العناصر المعابة هو 35% . بما أن العناصر المسحوبة قد سحبت بشكل عشوائى لذلك فإننا نجد :

$$P(ND) = P(N) \cdot P(D) = (0.65)(0.35) = 0.2275$$

وبشكل مشابه نحسب احتمالات بقية النقاط فى الجدول السابق فنجد أن :

X	0	1	2
P	0.4225	0.455	0.1225

تعريف (٣,١) متغير عشوائى حدانى Binomial Rondom Variable

إن عدد مرات النجاح X فى تجربة حدانية (مكونة من n اختبارا مكررا) يدعى بمتغير عشوائى حدانى .

نسمى توزيع الاحتمال للمتغير الحدانى بالتوزيع الحدانى ، وسنرمز له بالرمز

$b(x;n,p)$ ، لأن قيم هذا الاحتمال تتعلق بعدد الاختبارات المكررة وباحتمال النجاح فى كل اختبار . مثلا فى المثال السابق نجد أن :

$$P(X = 2) = f(2) = b(2;3,0.35) = 0.1225$$

لنعمم الآن المناقشة السابقة لنحصل على صيغة للتوزيع $b(x; n, p)$. إن ما نرغب به هو الحصول على صيغة تعطينا احتمال الحصول على x نجاح عند تكرار تجربة حدانية n مرة . لنحسب أولا احتمال الحصول على x نجاح و $n - x$ فشل فى ترتيب معين . بما أن الاختبارات المكررة مستقلة ، لذلك يمكن أن نحصل على الصيغة المطلوبة من حاصل ضرب كل الاحتمالات الموافقة للنتائج المختلفة . فإذا رمزنا لاحتمال النجاح بالرمز P ولاحتمال الفشل بالرمز $q = 1 - p$ فعندئذ يكون الاحتمال المطلوب فى ترتيب معين مساويا $p^x . q^{n-x}$. علينا الآن أن نحدد المجموع الكلى لعدد نقاط العينة فى تجربة حدانية فيها x نجاح و $n - x$ فشل ، وهذا العدد هو : $\binom{n}{x}$ لذلك فاحتمال الحصول على x نجاح فى التجربة الحدانية هو $\binom{n}{x} P^x . q^{n-x}$.

التوزيع الحدانى

إذا كان احتمال النجاح فى تجربة حدانية p ، واحتمال الفشل $q = 1 - p$ ، عندئذ يكون توزيع الاحتمال للمتغير العشوائى الحدانى الممثل لعدد مرات النجاح فى n اختبارا مستقلا هو :

$$b (x; n, p) = \binom{n}{x} P^x . q^{n-x} : x = 0, 1, 2, \dots , n$$

لاحظنا أنه إذا كان $n=2$, $P = 0.35$. فعندئذ يكون توزيع الاحتمال للمتغير X الممثل لعدد العناصر المعابة عند سحب عنصرين بشكل عشوائى من بضاعة مصنعة هو :

$$b (x; 2, 0.35) = \binom{2}{x} . (0.35)^x \, 0.65)^{2-x} : x = 0, 1, 2,$$

وهى نفس الصيغة المستخدمة فى الجدول السابق .

مثال (٣,٣)

بفرض أن احتمال أن بقاء مريض على قيد الحياة بعد إجراء عملية جراحية معينة هو 0.85 ، ما هو احتمال أن يبقى ثلاثة من خمسة مرضى آخرين على قيد الحياة عند إجراء نفس العملية لهم ؟

الحل

نلاحظ أن إجراء عملية لمريض يمثل اختباراً مستقلاً عن نتيجة أى عملية أخرى تجرى لمريض آخر . ثم إن $p = 0.85$ لكل اختبار ، لذلك :

$$b(3; 5, 0.85) = \binom{5}{3} (0.85)^3 (0.15)^2$$

$$= \frac{5!}{3!\,2!} (0.85)^3 (0.15)^2$$

$$= 0.138$$

ملاحظة

نلاحظ أن توزيع الاحتمال الحدانى يستمد اسمه من حقيقة أن الـ $n + 1$ عنصراً فى منشور نيوتن $(p + q)^n$ والتى توافق القيم $b(x; n, p) : x = 0, 1, \dots, n$.

$$(q + p)^n = \binom{n}{0} q^n + \binom{n}{1} p \cdot q^{n-1} + \binom{n}{2} p^2 \cdot q^{n-2} + \dots + \binom{n}{n} p^n$$

$$= b(0; n, p) + b(1; n, p) + b(2; n, p) \dots + b(n; n, p)$$

وبما أن $P + q = 1$ لذلك نلاحظ أن : $\sum_{x=0}^{n} b(x; n, p) = 1$ وهى الخاصة التى يجب أن يحققها أى توزيع احتمال .

من الملاحظ أيضا أننا نهتم على الغالب فى حساب $P[X < r]$ ، أو $P[a \leq x \leq b]$. ولحسن الحظ فإن جدول II يعطى قيم المجموع $B(r; n, p) = \sum_{x=0}^{n} b(x; n, p)$ ، وذلك من أجل $n = 5, 10, 15, 20$ ، ومن أجل أى قيمة لـ p من 0.1 إلى 0.9 .

مثال (٣,٤)

أطلق صياد خمس طلقات على هدف . فإذا علمت أن احتمال إصابته للهدف فى كل إطلاق هو 0.8 فما هو احتمال :

أ ــ أن يصيب الهدف مرتين تماما ؟
ب ــ أن يصيب الهدف مرتني على الأقل ؟
جـ ــ أن يصيب الهدف خمس مرات تماما ؟

الحل

نلاحظ أن عملية الإطلاق على الهدف تتم على شكل تكرارات (اختبارات) مستقلة بعضها عن بعض ، كما نلاحظ أن عملية الإطلاق هذه تمثل تجربة حدانية فيها $q = 0.2, p = 0.8, n = 5$. فبفرض أن عدد مرات إصابة الهدف هو X فإننا نجد أن :

$$P[X = 2] = b(2; 5, 0.8)$$

$$= \sum_{x=0}^{2} b(x; 5, 0.8) - \sum_{x=0}^{1} b(x; 5, 0.8)$$

وبالعودة إلى جدول II فى نهاية الكتاب نجد أنه من أجل $p = 0.8, r = 2, n = 5$

$$\sum_{x=0}^{2} b(x; 5, 0.8) = 0.0579$$

كذلك فإنه من أجل : $p = 0.8, r = 1, n = 5$

$$\sum_{x=0}^{1} b(x; 5, 0.8) = 0.0067$$

وحاصل الطرح بين القيمتين هو :

$$p[X = 2] = 0.0579 - 0.0067$$

$$= 0.0512$$

ولو حسبنا هذا الاحتمال بالطريقة العادية لوجدنا أن :

$$P[X = 2] = \binom{5}{2} (0.8)^2 . (0.2)^3$$

$$= \frac{5!}{2!3!} (0.8)^2 (0.2)^2$$

$$= 0.0512$$

وهى نفس النتيجة السابقة .

كذلك فإن :

$$P[X \geq 2] = 1 - P[X < 2]$$

$$= 1 - \sum_{x=0}^{1} b(x; 5, 0.8)$$

ومن الجدول II نجد أن المجموع السابق يساوى :

$$= 1 - 0.0067$$

$$= 0.9933$$

وأخيرا :

$$P[X = 5] = \sum_{x=0}^{5} b(x; 5, 0.8) - \sum_{x=0}^{4} b(x; 5, 0.8)$$

$$= 1 - 0.6723$$

$$= 0.3277$$

ونلاحظ أيضا أن :

$$P[X = 5] = \binom{5}{5} (0.8)^5 (0.2)^0$$

$$= (1) (0.8)^5 (1)$$

$$= 0.32768$$

وهى تقريبا نفس النتيجة السابقة .

مثال (٣,٥)

بفرض أن احتمال شفاء مريض فى عملية جراحية هو 0.9 . فما هو احتمال أن يشفى ثمانية مرضى من أصل عشرة أجريت لهم نفس العملية ؟

الحل

نلاحظ أن العمليات العشرة التى أجريت تمثل عشرة اختبارات فى تجربة حدانية فيها نتيجة كل اختبار ، إما شفاء للمريض أو عدمه ، كما نلاحظ أن احتمال شفاء المريض فى كل عملية (اختبار) ثابت من عملية إلى أخرى . لذلك فالتجربة المجراة هى تجربة حدانية فيها $x = 8, p = 0.9, n = 10$ ، وذلك بفرض أن عدد المرضى الذين تم شفاؤهم هو X . أما الاحتمال المطلوب فهو :

$$P[X = 8] = P[X \leq 8] - P[X \leq 7]$$

$$= \sum_{x=0}^{8} b(x; 10, 0.9) - \sum_{x=0}^{7} b(x; 10, 0.9)$$

وبالعودة إلى جدول II نجد أن :

$$P[X = 8] = 0.2639 - 0.0702$$
$$= 0.1937$$

نظرية (٣,٢)

إن توقع وتباين متغير عشوائى حدانى X يعطيان بالعلاقتين :

$$\mu = np, \sigma^2 = np.q$$

البرهان

لنرمز بـ X_i لعدد مرات النجاح فى الاختبار رقم i للتجربة الحدانية . نلاحظ أن

المتغيرات X_i مستقلة ، كما أن كل متغير من هذه المتغيرات يفترض إحدى القيمتين 1, 0 ، حسبما تكون نتيجة الاختبار فشل أو نجاح على الترتيب . كما نلاحظ أيضا أن مجموع عدد مرات النجاح هو :

$$X = X_1 + X_2 + \dots + X_n$$

والملاحظ أن التوقع الرياضى لكل متغير من المتغيرات X_i هو :

$$E(X_i) = 1.p + 0.q = p$$

ولذلك مهما تكن $i = 1,2, \dots, n$. وحسب النظرية (٣,,) نجد أن :

$$E(X) = E(X_1) + E(X_2) + \dots + (X_n)$$

$$= p + p + p + \dots + p$$

$$= np$$

أما تباين المتغير الحدانى X فيحسب كما يلى :

$$\sigma_X^2 = \sigma^2_{X_1 + X_2 + \dots + X_n}$$

وحسب النظرية (١١,٢) وعلى اعتبار أن المتغيرات $X_1 \dots , X_n$ مستقلة فإننا نجد أن :

$$\sigma_x^2 = \sigma^2_{x_1} + x_2 + x_3 + \dots + x_n$$

غير أن :

$$\sigma^2_{x_i} = E(X_i^2) - p^2$$

$$= (0)^2 . q + (1)^2 . p - p^2$$

$$\sigma^2_{x_i} = p.q$$

$$\sigma_x^2 = n\,p.q$$

وهو المطلوب برهانه

مثال (٣,٦)

احسب توقع وتباين عدد المرضى الذين تم شفاؤهم فى المثال (٣,٥) .

الحل

نلاحظ أن :

$$n = 10$$
$$p = 0.9$$
$$q = 0.1$$

باستخدام النظرية (٣,٢) نجد أن :

$$\mu = n.p = (10)(0.9) = 9$$
$$\sigma^2 = n.p.q = (10)(0.9)(0.1) = 0.9$$

لنفرض الآن لكل اختبار (فى التجربة الحدانية) أكثر من نتيجتين ممكنتين ، فعندئذ نقول بأن لدينا تجربة متعددة الحدود (multinoumial experiment) . فلو اعتبرنا أن ورق اللعب مؤلف من أربعة أنواع بحسب النوع (ديناری ـــ بستونی ـــ كبه ـــ سباتی) فعند سحب ورقة من ورق اللعب مع الإعادة بصورة عشوائية نكون أمام تجربة متعددة الحدود ، لأن نتيجة السحب ستكون واحدة من أربعة أنواع .

وبشكل عام ، إذا كانت نتيجة التجربة إحدى النتائج الـ k الممكنة $E_1, E_2, \ldots, E_k$ ، وإذا كانت الاحتمالات الموافقة لهذه النتائج $P_1, \ldots, P_k$ ، فعندئذ يمثل التوزيع المتعدد الحدود احتمال الحصول على النتائج $E_1, E_2, \ldots, E_k$ عدداً من المرات $X_1, \ldots, X_k$ على الترتيب فى n تكراراً مستقلاً ، حيث إن $X_1, X_2 \ldots, X_n = n$.

سنرمز لتوزيع الاحتمال هذا بالرمز $f(x_1, \ldots, x_k, P_1, \ldots, P_k, n)$ من الواضح أن $P_1 + \ldots + P_k = 1$ ، لأن نتيجة كل اختبار (أو تكرار) تمثل إحدى النتائج الـ k الممكنة .

وللحصول على الصيغة العامة لهذا التوزيع ، سنتبع نفس الخطوات التى سرنا عليها لدى الحصول على التوزيع الحدانى . فنفرض ترتيب معين للنتائج ، ثم نحسب احتمال هذا الترتيب فنجد أنه : $P_1^{x_1} . P_2^{x_2} \ldots P_k^{x_k}$

ثم نبحث عن المجموع العام لهذه الترتيبات ، وهذا العدد يمثل عدد الطرق التى يمكن أن

تجزأ بها مجموعة مؤلفة من n عنصراً إلى عدد من الخلايا k تحوى الأولى على x_1 عنصراً ، والثانية على x_k عنصراً . وهذا العدد من التجزئات يمكن أن يتم بعدد من الطرق مساو لـ :

$$\binom{n}{x_1, x_2, \ldots, x_k} = \frac{n!}{x_1!\, x_2! \ldots x_k!}$$

وبما أن كل التجزئات متنافية تبادلياً (أى لا يمكن أن نحصل على تجزئتين مختلفتين فى وقت واحد) ، واحتمال وقوع أى منها واحد ويساوى $P_1^{x_1} \ldots P_k^{x_k}$ ، لذلك فإن الاحتمال المطلوب هو :

$$f(x_1, \ldots, x_k ; p_1, \ldots, p_k, n) = \frac{n!}{x_1! \ldots x_k!} p_1^{x_1} \ldots p_k^{x_k}$$

التوزيع المتعدد الحدود Multinomial distribution

إذا كانت النتائج الممكنة لتجربة معينة هى $E_1, E_2 \ldots, E_k$ ، واحتمالاتها الموافقة هى $P_1, \ldots, P_k$ فعندئذ يكون توزيع الاحتمال للمتغيرات $X_1, \ldots, X_k$ الممثلة لعدد المرات التى ستقع فيها النتائج $E_1, \ldots, E_k$ على الترتيب فى n تجربة مستقلة معطى بالعلاقة السابقة .

وحيث إن : $\sum_{i=1}^{k} x_i = n, \sum_{i=1}^{k} P_i = 1$

ومما يجدر ذكره أن التوزيع المتعدد قد اشتق اسمه من حقيقة أن المنشور المتعدد $(P_1 + P_2 + \ldots + P_k)^n$ يوافق جميع قيم $f(x_1, \ldots, x_k; P_1, \ldots P_k, n)$.

مثال (٣,٧)

لنحسب احتمال الحصول على مجموع يساوى 3 أو 9 مرتين ، ثنائية متشابهة مرة واحدة ، وأى شكل آخر ثلاث مرات لدى إلقاء حجري نرد ست مرات متتالية .

الحل :

لنشكل الحوادث التالية :

$A_1 =$ { الحصول على مجموع 3 أو 9 }
$A_2 =$ { الحصول على ثنائية متشابهة }
$A_3 =$ { الحصول على أى شكل آخر }

والمراد الآن أن تقع الحوادث A_1 مرتين ، A_2 مرة واحدة ، A_3 ثلاث مرات .

نلاحظ أن :

$$P_1 = P(A_1) = \frac{4+2}{36} = \frac{1}{36}, \; P_2 = P(A_2) = \frac{6}{36} = \frac{1}{6}$$

$$P_3 = P(A_3) = 1 - \left[\frac{1}{36} + \frac{6}{36}\right] = 1 - \frac{7}{36} = \frac{29}{36}$$

لذلك فالاحتمال المطلوب هو :

$$f\left(2, 1, 3; \frac{1}{36}, \frac{1}{6}, \frac{29}{36}, 6\right) = \frac{6!}{2!\,.\,1!\,3!}\left(\frac{1}{36}\right)^2 \cdot \left(\frac{1}{6}\right)^1 \cdot \left(\frac{29}{36}\right)^3$$

(٣,٤) التوزيع الهندسى الزائدى Hypergeometric distribution

من الملاحظ أنه لا يمكن تطبيق التوزيع الحدانى فى حساب احتمال الحصول على أربع أوراق حمراء ، وذلك عند سحب ست أوراق من ورق اللعب بصورة عشوائية دفعة واحدة أو بالتتالي ولكن بدون إعادة إلا أنه إذا سحبنا ورقة بصورة عشوائية وأعدناها الى مجموعة ورق اللعب ثم خلطنا ورق اللعب خلطا جيدا ، وكررنا العملية السابقة مثلا ست مرات ، فإنه يمكن عندئذ حساب الاحتمال السابق المذكور . بشكل عام يمكن حساب الاحتمال السابق عند السحب بدون إعادة على النحو التالى :

نفرض الحوادث :

$A =$ { ظهور أربع أوراق حمراء }
$B =$ { ظهور ورقتين سوداوين }

نلاحظ أن فضاء العينة S فى هذه التجربة يتألف من $\binom{52}{6}$ نقطة ، وبين هذه النقاط عددا يوافق الحادث A هو $\binom{26}{4}$. كما أنه يوافق الحادث B عددا من النقاط مساوٍ لـ $\binom{26}{2}$ ، والاحتمال المطلوب يعطى بالعلاقة :

$$P = \frac{\binom{26}{4}\binom{26}{2}}{\binom{52}{6}} = 0.23$$

نسمى تجربة سحب n شيئا فى مجموعة من الأشياء (مؤلفة من N شيئا منها K شيئا مكتوب على كل منها نجاح و N - K فشلا) وظهور x نجاحا و n - x فشلا بتجربة هندسية زائدية . من الملاحظ أن التجربة الهندسية الزائدية تتصف بالخواص التالية :

(١) سحب n عنصرا من N شيئا .

(٢) السحب يتم دفعة واحدة وبدون إعادة .

(٣) K شيئا من الـ N المفروضة مكتوب على كل منها نجاح و N - K فشل .

تعريف (٣,٢) التوزيع الهندسى الزائدي

نسمى عدد النجاحات فى تجربة هندسية زائدية بالمتغير الهندسى الزائدى ، كما أن احتمال هذا المغير يدعى بالتوزيع الهندسى الزائد ، ونظرا لكون هذا الاحتمال معتمدا على K, N فإننا سنرمز له بالرمز $h(x; N, n, k)$.

مثال (٣,٨)

اختيرت لجنة مؤلفة من أربعة أشخاص بشكل عشوائى من مجموعة أربعة كيميائيين ، وسبعة فيزيائيين . ما هو توزيع الاحتمال لعدد الكيميائيين فى اللجنة ؟

الحل

نلاحظ أن التجربة المجراة هى تجربة هندسية زائدية . لنرمز بـ X لعدد الكيميائيين الموجودين فى اللجنة . نجد أن قيم المتغير X هى 0, 1, 2, 3, 4 ، وتوزيع الاحتمال للمتغير الهندسى الزائدى x هو :

$$h(x;\ 11,\ 4,7) = P\,[\,X = x\,]$$

كملا نلاحظ أن :

$$h\,(0;\ 11,\ 4,\ 7) = \frac{\binom{4}{0}\ \binom{7}{4}}{\binom{11}{4}} = 0.106$$

$$h\,(1;\ 11,\ 4,\ 7) = \frac{\binom{4}{1} \cdot \binom{7}{3}}{\binom{11}{4}} = 0.424$$

$$h\,(2;\ 11,\ 4,\ 7) = \frac{\binom{4}{2}\ \binom{7}{2}}{\binom{11}{4}} = 0.38$$

$$h\,(3;\ 11,\ 4,\ 7) = \frac{\binom{4}{3}\ \binom{7}{1}}{\binom{11}{4}} = 0.0848$$

$$h\,(4;\ 11,\ 4,\ 7) = \frac{\binom{4}{4}\ \binom{7}{0}}{\binom{11}{4}} = 0.0030$$

وهكذا نجد أن جدول توزيع X هو :

X	0	1	2	3	4
P	0.106	0.424	0.38	0.0848	0.00303

ومن السهل ملاحظة أن توزيع الاحتمال يمكن الحصول عليه من الصيغة التالية :

$$h(x; 11, 4, 7) = \frac{\binom{4}{x} \cdot \binom{7}{4-x}}{\binom{11}{4}} \quad : x = 1, 0, 2, 3, 4$$

لنعمم الآن المثال (٣,٨) فى البحث عن احتمال سحب n عنصرا من ضمن N شيئا وظهور x نجاح ، و n - x فشل . أن عدد إمكانيات اختيار n شيئا (فى المجموعة التى تحوى N عنصرا) وبشكل عشوائى هو $\binom{N}{n}$ طريقة . هذا العدد يمثل عدد نقاط فضاء العينة S . سنفرض أن هذه النقاط متساوية فى إمكانية وقوعها .

نلاحظ أن هناك $\binom{K}{x}$ طريقة لاختيار x نجاحا من ضمن K نجاح ، وكل طريقة من هذه الطرق توافق ($\binom{N-K}{n-x}$) طريقة لاختيار n - x فشلا من ضمن الـ N - K فشل (والموجودة ضمن الـ N شيئا) ، وهكذا نجد أنه يوافق الحادث المراد حساب احتماله عددا من النقاط ($\binom{N-K}{n-x}$) . ($\binom{K}{x}$) ، إذن فالاحتمال المطلوب والذى نرمز له بالرمز h(x; N, n, k) يعطى من خلال التعريف التالى :

التوزيع الهندسى الزائدى

إن التوزيع الاحتمالى للمتغير العشوائى الهندسى الزائدى X (الممثل لعدد مرات النجاح عند سحب n عنصرا بشكل عشوائى من N عنصرا تحوى K نجاحا و N - K فشلا) يعطى بالعلاقة التالية :

$$h(x; N, n, k) = \frac{\binom{k}{x} \cdot \binom{n-k}{n-x}}{\binom{N}{n}} \quad , x = 0, 1, 2, \dots, n$$

نظرية (٣,٣)

إن توقع وتباين متغير عشوائى هندسى زائدى بالتوزيع h (x; N, n, k) يعطيان بالعلاقتين التاليتين :

$$\mu = \frac{nK}{N} \quad , \sigma^2 = \frac{(N-n)}{(N-1)} .n. \frac{K}{N} . (1 - \frac{K}{N})$$

البرهان

للبحث عن التوقع الرياضى للمتغير الهندسى الزائدى نكتب ما يلى :

$$\mu = E(X) = \sum_{x=0}^{n} x \frac{\binom{K}{x} \cdot \binom{N-K}{n-x}}{\binom{N}{n}}$$

$$= K \sum_{x=1}^{n} \frac{(K-1)!}{(x-1)!\,(K-x)!} , \frac{\binom{N-K}{n\ \ x}}{\binom{N}{n}}$$

$$= K \cdot \sum_{x=1}^{n} \frac{\binom{K-1}{x-1}\binom{N-K}{n-x}}{\binom{N}{n}} \qquad Y = x - 1$$

وبوضع $Y = x - 1$ فإننا نجد أن :

$$\mu = K \cdot \sum_{y=0}^{n-1} \frac{\binom{K-1}{y}\binom{N-K}{n-1-y}}{\binom{N}{n}}$$

$$= K \cdot \sum_{y=0}^{n-1} \frac{\binom{K-1}{y}\binom{(N-1)-(K-1)}{(n-1)-\ \ y}}{\binom{N}{n} \cdot \binom{N-1}{n-1}}$$

$$= \frac{K n}{N}$$

لأن المجموع الأخير يمثل مجموع كل الاحتمالات فى التجربة الهندسية الزائدية ، عندما نختار (n - 1) عنصرا من (N - 1) عنصر تحتوى على K - 1 نجاحا . وهذا المجموع يساوى الواحد حسب خواص الكثافة الاحتمالية .

لإيجاد تباين التوزيع الهندسى الزائدى ، نتبع نفس الخطوات السابقة للحصول على :

$$E[X(X-1)] = \frac{K.(k-1)\,n\,(n-1)}{N.(N-1)}$$

واعتمادا على النظرية (٢,٦) نجد أن :

$$\sigma^2 = E(X^2) - \mu^2$$

$$= E[X(X-1)] + \mu - \mu^2$$

$$= \frac{K.(K-1)\,n\,(n-1)}{N.(N-1)} + \frac{nK}{N} - \left(\frac{nK}{N}\right)^2$$

$$= \frac{nK.(N-K)(N-n)}{N^2.(N-1)}$$

$$= \frac{(N-n)}{(N-1)} . n . \frac{K}{N} . \left(1-\frac{K}{N}\right)$$

مثال (٣,٩)

تحتوى كل رزمة من بضاعة مصنعة على 40 عنصراً ، وتقبل رزمة ما إذا حوت عددا من العناصر المعيبة لا يتجاوز الثلاثة عناصر . لفحص أى رزمة نأخذ منها عينة مؤلفة من خمسة عناصر ونفحصها ، فإذا وجدنا فيها عنصرا معيباً فإننا نرفض الرزمة . ما هو احتمال وجود عنصر معيب فى عينة عند فحص رزمة . إذا علمت أن الرزمة المفحوصة تحتوى على ثلاثة عناصر معيبة ؟

الحل

باستخدام التوزيع الهندسى الزائدى من أجل $n = 5, N = 40, K = 3, x = 1$ ، فإننا نجد أن احتمال وجود عنصر معيب فى العينة هو :

$$h(1; 40, 5, 3) = \frac{\binom{3}{1}\binom{37}{4}}{\binom{40}{5}} = 0.3011$$

مثال (٣,١٠)

باستخدام نظرية تشبيشيف ، لنبحث عن المجال $(\mu - 2\sigma, \mu + 2\sigma)$ فى المثال السابق .

الحل

اعتمادا على النظرية (٣,٣) نجد أن :

$$\mu = \frac{(5) . (3)}{40} = \frac{3}{8} = 0.375$$

$$\sigma^2 = \frac{(40-5)}{(39)} (5) . (\frac{3}{40}) . (1 - \frac{3}{40}) = 0.3113$$

ويأخذ الجذر التربيعى للمقدار الأخير ، فإننا نجد أن $\sigma = 0.558$ ، والمجال المطلوب هو :

$$(- 0.741, 1.491)$$

وحسب نظرية تشبيشيف نجد أن احتمال وقوع عدد العناصر المعيبة (التى نجدها فى عينة مؤلفة من خمسة عناصر مأخوذة من رزمة مؤلفة من 40 عنصرا فيها ثلاثة عناصر معيبة) فى المجال (0.741, 1.491 -) على الأقل $\frac{3}{4}$. وهذا يعنى أنه فى ثلاثة أرباع الفترة الزمنية ستحتوى العينة المؤلفة من خمسة عناصر على عنصرين معيبين على الأقل .

إذا كان n صغيرا بالنسبة لـ N فإننا نجد أن الاحتمال فى كل سحبة سيتغير بشكل تافه . وهذا يعنى أن تجربتنا ستأخذ شكل التجربة الحدانية وإذ يمكن تقريب التوزيع الهندسي الزائدي في هذه الحالة إلى التوزيع الحدانى باستخدام $P = \frac{K}{N}$. ويقرب كل من المتوقع والتباين بوساطة العلاقات :

$$\mu = n.p. = \frac{n . K}{N}$$

$$\sigma^2 = n.p.q. = n. \frac{K}{N} . (1 - \frac{K}{N})$$

وبمقارنة هذه الصيغ مع ما جاء فى النظرية (٣,٣) نلاحظ أن التوقع الرياضى ظل نفسه ، بينما اختلف التباين بتعديل العامل $\frac{N-n}{N-1}$ ، هذا تافة عندما تكون n صغيرة بالنسبة لـ N .

مثال (٣,١١)

أعلنت شركة لصنع الإطارات المطاطية أن شحنتها المؤلفة من 3200 إطار والمتجهة إلى أسواق البيع تحتوى على 800 إطار معيب . لنحسب احتمال أن يشترى شخص ما عشرة إطارات من هذه الكمية بشكل عشوائى فيجد بينها إطارين معيبين .

الحل

بما أن العدد N = 3200 كبير بالنسبة لحجم العينة المختارة n = 10 ، لذلك فإننا سنقوم بتقريب الاحتمال المطلوب باستخدام التوزيع الحدانى . أن احتمال الحصول على إطار معيب يساوى $0.25 = \frac{800}{3200}$. لذلك فإن احتمال الحصول على إطارين معيبين يحسب بالعلاقة :

$$h(2; 3200, 10, 800 \simeq b(2; 10, 0.25)$$

$$= \sum_{x=0}^{2} b(x; 10, 0.25) - \sum_{x=0}^{1} b(x; 10, 0.25)$$

وباستخدام الجدول II نجد أن :

$$= 0.5256 - 0.244$$

$$= 0.2816$$

يمكن تعميم التوزيع الهندسى الزائدى لمعالجة الحالة التالية :

نفرض أن المجموعة المؤلفة من N عنصرا تحتوى على k خلية $A_1, A_2, \ldots, A_k$ ، حيث تحتوى هذه الخلايا على الترتيب $a_1, a_2, \ldots, a_k$ عنصرا . لنحسب احتمال احتواء عينة حجمها n مختارة بصورة عشوائية من الـ N عنصرا على x_1 عنصرا من الخلية A_1 . وعلى x_2

عنصرا من الخلية الثانية A_2 وهكذا ... و x_k عنصرا من الخلية A_k بحيث يكون :

$$\sum_{i=1}^{k} x_i = n$$

هذا الاحتمال سنرمز له بالرمز :

$$f(x_1, \ldots, x_k; a_1, a_2, \ldots, a_k, N, n)$$

للحصول على الصيغة العامة لهذا الاحتمال ، نلاحظ أن عدد الطرق الممكنة لتحقيق هذه التجربة هو $\binom{N}{n}$. كما نلاحظ أن هناك $\binom{a_1}{x_1}$ طريقة لاختيار x_1 عنصرا من عناصر A_1 ، ومن أجل كل طريقة من هذه الطرق هناك $\binom{a_2}{x_2}$ طريقة لاختيار x_2 عنصرا من عناصر A_2 ، ولذلك فإنه يمكن أن نختار x_1 عنصرا من A_1 و x_2 عنصرا من A_2 بعدد من الطرق مساوٍ لـ $\binom{a_1}{x_1}\binom{a_2}{x_2}$ طريقة . وبمتابعة هذه العملية نجد أننا نستطيع أن نختار الـ n عنصرا ، والتى تحتوى على x_1 من A_1 و x_2 من A_2 , ... , x_k من A_k بعدد من الطرق مساوٍ لـ $\binom{a_1}{x_1}\binom{a_2}{x_2}\cdots\binom{a_k}{x_k}$ طريقة .

والاحتمال المطلوب يعرف كما يلى :

تعميم التوزيع الهندسى الزائدى

إذا احتوت مجموعة مؤلفة من N عنصرا على K خلية $A_1, \ldots, A_k$ بالعناصر :

$a_1, \ldots, a_k$ على الترتيب . يكون توزيع الاحتمال للمتغيرات العشوائية $X_1, \ldots, X_k$ الممثلة لعدد العناصر المختارة من هذه الخلايا بشكل عشوائى من خلال عينة حجمها n هو :

$$f(x_1, \cdots, x_k; a_1, \cdots, a_k, N, n) = \frac{\binom{a_1}{x_1}\binom{a_2}{x_2}\cdots\binom{a_k}{x_k}}{\binom{N}{n}}$$

حيث :

$$\sum_{i=1}^{k} a_i = N \quad , \quad \sum_{i=1}^{k} x_i = n$$

مثال (٣,١٢)

سحبنا ، وبشكل عشوائى ، خمسة أوراق من ورق اللعب المؤلف من 52 ورقة . ما هو احتمال أن يكون عدد أوراق الديناري المسحوبة 2 والبستونى 1 والكبة 2 ؟

الحل

لنفرض أن :

$X_1 =$ { عدد أوراق الديناري المسحوبة ضمن الخمس أوراق }
$X_2 =$ { عدد أوراق البستوني المسحوبة ضمن الخمس أوراق }
$X_3 =$ { عدد أوراق الكبة المسحوبة ضمن الخمس أوراق }
$X_4 =$ { عدد أوراق سباتي المسحوبة ضمن الخمس أوراق }

نلاحظ أن الورق المؤلف من 52 ورقة يحتوى على أربع خلايا كل خلية تحتوى على 13 ورقة ، هذه الخلايا هى الديناري ، البستوني ، الكبة والسباتي .

والاحتمال المطلوب هو :

$$f(2, 1, 2, 0, 13, 13, 13, 52, 5) = \frac{\binom{13}{2}\binom{13}{1}\binom{13}{2}\binom{13}{0}}{\binom{52}{5}} = 0.03$$

(٣,٥) التوزيع البواسوني Poisson distribution

للتوزيع البواسوني تطبيقات واسعة . فهو يقدم بصورة خاصة نموذجا جيدا للمعلومات التى تأخذ شكل التعداد ، حيث يمثل X فيه عدد الحوادث النادرة (الملاحظة) فى وحدة معينة زمانا كانت ، أم مسافة ، أم مساحة ، أم حجما . وكأمثلة على هذا العدد ، نسوق مايلى :

(١) عدد المكالمات الهاتفية المتبادلة بين الساعة 5 + t, 5 بين نيويورك وجدة .
(٢) عدد الذرات الصادرة فى ميكروثانية عن كمية من مادة مشعة .
(٣) عدد الإلكترونات التى يصدرها مهبط مسخن فى فترة زمنية محددة .
(٤) عدد حوادث السير خلال زمن محدد .
(٥) عدد الأخطاء المطبعية فى صفحة ما .
(٦) عدد باكيتات السجاير المبيعة يوم الاثنين فى دكان بقالة .
(٧) عدد البكتريا الموجودة فى حجم صغير من سائل معين .
(٨) عدد ذرات الغاز فى منطقة جزئية v من وعاء حجمه V .

توضح الأمثلة السابقة مدى تنوع واتساع تطبيقات التوزيع البواسونى .

إن التجربة التى تقدم لنا قيما عددية لمثل هذه المتغيرات نطلق عليها اسم تجربة بواسونية ، والملاحظ أن التجربة البواسونية تتمتع بالخواص التالية :

(١) أن عدد النجاحات التي نحصل عليها في فترة زمنية (مثلا عدد المكالمات الهاتفية في فترة زمنية) مستقل عن أى عدد لهذه النجاحات في فترات أخرى .

(٢) أن احتمال الحصول على نجاح مفرد في زمن قصير يتناسب مع طول هذا المجال الزمني القصير ولايعتمد هذا الاحتمال على عدد النجاحات التي نحصل عليها في مجالات زمنية أخرى .

(٣) احتمال الحصول على أكثر من نجاح (في مثل هذا المجال الزمني القصير) مهمل .

ملاحظة : يمكن أن تكون الفترة الزمنية المفروضة منطقة معينة (طول — مساحة — حجم) .

تعريف (٣،٣) متغير عشوائى بواسونى **Poisson random variable**

إن عدد النجاحات X (مثلا عدد المكالمات الهاتفية) فى تجربة يدعى بمتغير عشوائى بواسونى ، كما يدعى توزيع الاحتمال للمتغير X بالتوزيع البواسونى . سنرمز لهذا النوع بالرمز $P(x; \mu)$ ، ولقد استخدما الرمز السابق عوضا عن $f(x)$ لكى نشير إلى أن دالة الكثافة الاحتمالية تتعلق بمتوسط عدد النجاحات X خلال فترة زمنية أو منطقة معينة .

لنفرض فى التوزيع الحدانى أن n كبيرة جدا و p صغيرة جدا ، بحيث إن الجداء $np = \mu$ يبقى مساويا لعدد ثابت . إذا عوضنا عن p بـ $\frac{\mu}{n}$ فى عبارة التوزيع الحدانى ، فإننا نجد أن :

$$f(x) = \frac{n!}{x!(n-x)!} P^x \cdot q^{n-x}$$

$$= \frac{n!}{x!(n-x)!} \left(\frac{\mu}{n}\right)^x \cdot \left(1 - \frac{\mu}{n}\right)^{n-x}$$

$$= \frac{n}{n} \cdot \frac{n-1}{n} \cdot \frac{n-2}{n} \cdots \frac{n-x+1}{n} \cdot \frac{\mu^x (1-\frac{\mu}{n})^n \cdot (1-\frac{\mu}{n})^{-x}}{x!}$$

ونلاحظ أن النسب : $\frac{n-1}{n} , \frac{n-2}{n} , \ldots , \frac{n-x+1}{n}$ قريبة جدا من الواحد لأن n كبيرة جدا ، كما فرضنا ، وقيمة x ثابتة بالنسبة لـ n . وبأخذ نهاية طرفى العلاقة السابقة نجد أن الطرف الأيمن سيأخذ الشكل التالى :

$$\frac{\mu^x}{x!} \cdot (1-\frac{\mu}{n})^n$$

ويبرهن فى الرياضيات أنه عندما تصبح n كبيرة جدا فإن الكمية $(1+\frac{1}{n})^n$ تقترب من عدد نرمز له بالرمز e (حيث إن e = 2.7183) ، ونعبر عن ذلك بقولنا إن $(1+\frac{1}{n})^n \underset{n\to\infty}{\longrightarrow} e$ كما ويمكن البرهان على أن . $(1-\frac{\mu}{n})^n \underset{n\to\infty}{\longrightarrow} e^{-\mu}$ ، وهكذا نجد أن الشكل الحدى لعبارة f(x) عندما تسعى n اللانهاية هو :

$$P(x;\mu) = \frac{\mu^x}{x!} e^{-\mu} , \quad x = 0, 1, 2, \cdots , \mu > 0$$

ومن المتوقع أن تعطى هذه العبارة نفس الاحتمالات التى تعطيها عبارة التوزيع الحدانى تقريبا ، وذلك شريطة أن تكون n كبيرة ، n.p صغيرة نسبيا . هذا ويمكن من أن $P(x;\mu)$ تحقق شرطى دالة الكثافة الاحتمالية .

مثال (٣,١٣)

تنتج آلة نوعا معينا من العناصر . إذا علم أن كل 1000 قطعة من إنتاج هذه الآلة تحتوى فى المتوسط على قطعة واحدة معيبة . فاحسب احتمال أن تحتوى عينة من إنتاج هذه الآلة مؤلفة من 8000 قطعة على أقل من 7 قطع معيبة .

الحل

لنرمز لعدد القطع المعيبة بين الـ 8000 قطعة بالرمز x ، فيكون المطلوب :

$$P[X < 7] = \sum_{x=0}^{6} b(x; 8000, 0.001)$$

$$= \sum_{x=0}^{6} P(x; 8)$$

$$= 0.3134$$

والملاحظ عند حل هذا المثال أن التجربة الحدانية $P = 0.001, n = 8000$. وبما أن P صغيرة جدا و n كبيرة . لذلك قربنا التوزيع الحدانى إلى البواسونى باستخدام $n.p = (8000)(0.001) = 8$

التوزيع البواسونى

إن توزيع الاحتمال للمتغير العشوائى البواسونى X والممثل لعدد مرات النجاح التى نحصل عليها خلال فترة زمنية محددة أو منطقة معينة هو :

$$P(x; \mu) = e^{-\mu} : \frac{\mu^x}{x!} : x = 0, 1, 2, \ldots$$

حيث يمثل μ متوسط عدد مرات النجاح خلال نفس الفترة $e = 2.7183$ ، يعطينا الجدول III مجموع الاحتمالات البواسونية $\sum_{x=0}^{r} P(x; \mu)$ ، وذلك من أجل بعض قيم من $\mu = 0.1$ إلى $\mu = 18$.

مثال (٣,١٤)

يتسلم مقسم هاتف المخابرات الخارجية بين الساعة العاشرة والثانية عشرة بمعدل مخابرتين فى الدقيقة .

(١) لنحسب احتمال ألا يتسلم هذا المقسم أية مخابرة خارجية خلال دقيقة معينة ؟

(٢) ثم لنحسب احتمال أن يتسلم مخابرتين خلال دقيقة ؟

الحل

لنرمز لعدد المكالمات الخارجية خلال دقيقة بالرمز X . نلاحظ أن احتمال ألا يتسلم هذا المقسم أية مخابرة خارجية خلال دقيقة معينة هو :

$$P(X = 0) = \bar{e}^{\mu} \cdot \frac{\mu^0}{0!} = \bar{e}^{\mu}$$

غير أن معدل المخابرات الخارجية خلال دقيقة $\mu = 2$. لذلك فإن :

$$P(X = 0; 2) = e^{-2} = 0.135$$

كما أن احتمال أن يتسلم مخابرتين خلال دقيقة معينة هو :

$$P(X = 2) = P(2 ; 2) = \bar{e}^{2} \cdot \frac{2^2}{2!} = 0.27$$

نظرية (٣,٤)

إن متوسط وتباين التوزيع البواسونى $P(x; \mu)$ يساوى μ .

البرهان

$$E(X) = \sum_{x=0}^{\infty} x . P(x, \mu)$$

$$= \sum_{x=0}^{\infty} x \cdot \bar{e}^{x} \cdot \frac{\mu^x}{x!}$$

$$= \mu \sum_{x=1}^{\infty} \frac{e^{-\mu} \mu^{x-1}}{(x-1)!}$$

وبفرض أن $y = x - 1$ نجد :

$$= \mu \cdot \sum_{y=0}^{\infty} e^{-\mu} \frac{\cdot \mu^y}{y!}$$

$$= \mu$$

لأن :

$$\sum_{y=0}^{\infty} e^{-\mu} \frac{\mu^y}{y!} = \sum_{y=0}^{\infty} P(y ; \mu) = 1$$

لنبحث عن :

$$E[X(X-1)] = \sum_{x=0}^{\infty} x(x-1) \cdot \bar{e}^{\mu} \frac{\mu^x}{x!}$$

$$= \mu^2 \cdot \sum_{x=2}^{\infty} \frac{\bar{e}^{\mu} \cdot \mu^{x-2}}{(x-2)!}$$

فإذا فرضنا أن : $y = x - 2$ فإننا نجد :

$$= \mu^2 \cdot \sum_{y=0}^{\infty} \bar{e}^{\mu} \frac{\mu^y}{y!}$$

$$= \mu^2$$

لذلك فإن :

$$\sigma^2 = E[X(X-1)] + \mu - \mu^2 = \mu^2 + \mu - \mu^2 = \mu$$

نلاحظ فى المثال (١٤و٣) أن $\mu = 2$ ، وأن $\sigma^2 = 2$ أيضا ، لذلك فإن
نستطيع باستخدام متباينة تشبيشيف أن نتأكد من أن متغيرنا العشوائى سيقع فى المجال
باحتمال على الأقل $\frac{3}{4}$. لذلك يمكن القول بأنه خلال ثلاث أرباع الدقيقة
سيستقبل المقسم عددا من المكالمات الهاتفية من 0.838 إلى 4.838 مكالمة .

(٣,٦) التوزيع الحدانى السالب Negative binomial distribution

نفرض أن تجربتنا تتمتع بنفس خواص التجربة الحدانية مع إضافة أننا نكرر التجربة حتى نحصل على عدد معين من النجاحات . لذلك بدلا من البحث عن احتمال الحصول على x نجاح خلال الـ n اختبارا (حيث n ثابت) ، فإننا سنبحث عن احتمال الحصول على النجاح ذى الرقم k فى الاختبار ذى الرقم x . هذا النوع من التجارب يدعى بالتجارب الحدانية السالبة .

لنحسب مثلا احتمال الحصول على ثالث صورة فى سابع قذفة لقطعة نقود متوازنة ومتماثلة . نلاحظ أن هذا الاحتمال هو جداء احتمالين : أولهما احتمال الحصول على صورتين خلال الست قذفات الأولى ، وهذا الاحتمال يساوى :

$$\binom{6}{2} p^2 . q^4$$

والاحتمال الثانى هو احتمال الحصول على وجه الصورة فى القذفة السابقة ويساوى p .
وجداء الاحتمالين هو :

$$\binom{6}{2} p^3 . q^4$$

فإذا عوضنا عن $p = \frac{1}{2}$, $q = 1 - p$ ، لوجدنا أن الاحتمال المطلوب يساوى :

$$\binom{6}{2} (\frac{1}{2})^3 . (\frac{1}{2})^4 = 0.11718$$

تعريف (٣,٤) المتغير العشوائي الحداني السالب

Negative Binomial Random Variable

إن عدد الاختبارات X التى تعطى k نجاحا فى التجربة الحدانية السالبة يدعى بالمتغير العشوائى الحدانى السالب .

كما أن توزيع الاحتمال للمتغير الحدانى السالب X يدعى بالتوزيع الحدانى السالب ، وسنرمز له بالرمز $b^*(x; k, p)$ لأن قيمة هذا التوزيع تتعلق بعدد النجاحات المطلوبة ، وباحتمال النجاح فى كل اختبار . للحصول على الصيغة العامة للاحتمال $b^*(x; k, p)$ نحسب احتمال الحصول على نجاح فى الاختبار ذى الرقم x بعد حصولنا على k - 1 نجاحا و x - k فشلا فى ترتيب معين . بما أن الاختبارات مستقلة فإنه بإمكاننا ضرب كل الاحتمالات الموافقة لكل نتيجة مطلوبة . وبما أن احتمال الحصول على نجاح فى كل إختبار هو p وعلى فشل هو $q = 1 - p$ لذلك فاحتمال الحصول على نجاح فى الاختبار ذى الرقم x وذلك فى الترتيب المعين $p^{k-1} . q^{x-k} . p = p^k . q^{x-k}$ ، ويوجد عدد من التراتيب السابقه يساوى $\binom{x-1}{k-1}$ ، ولذلك فالاحتمال المطلوب والممثل لاحتمال الحصول على النجاح ذى الرقم k فى الاختبار ذى الرقم x يساوى :

$$b^* (x; k, p) = \binom{x-1}{k-1} p^k . q^{x-k} : x = k, k + 1, \ldots$$

التوزيع الحدانى السالب

إذا كان احتمال الحصول على نجاح فى كل اختبار هو p ، وعلى فشل هو $q = 1 - p$ ، فإن توزيع الاحتمال لعدد الاختبارات المجراة بقصد الحصول على k نجاح و $x - k$ فشل يعطى بالعبارة السابقة .

مثال (٣,١٥)

ما هو احتمال الحصول على صورتين أو كتابتين مرتين عند إلقاء ثلاث قطع نقود سبع مرات ؟

الحل

باستخدام التوزيع الحدانى السالب من أجل $x = 7$, $K = 2$, $p = \frac{3}{4}$

فإننا نجد أن :

$$b^* (7; 2, \frac{3}{4}) = \binom{6}{1} (\frac{3}{4})^2 (\frac{1}{4})^5$$
$$= 0.00109$$

لقد اشتق التوزيع الحدانى السالب اسمه من حقيقة أن كل عنصر فى منشور $\underline{k}(1 - q)^{-k}$ يوافق قيمة من قيم $b^*(x; k, p)$ من أجل $x = k, k + 1, k + 2$ لندرس الحالة الخاصة للتوزيع الحدانى السالب من أجل $k = 1$ ، عندئذ يمثل التوزيع الحدانى السالب توزيع الاحتمال لعدد الاختبارات المجراة والموافقة لنجاح وحيد . فمثلا فى تجربة إلقاء قطعة نقود متوازنة ومتماثلة ، فإننا سنلقى قطعة النقود حتى يظهر وجه الصورة (نجاح) لأول مرة ، وعندها نتوقف عن قذف قطعة النقود . سنسحب مثلا احتمال الحصول على وجه صورة فى القذفة الرابعة . وفى هذه الحالة سيأخذ التوزيع الحدانى السالب الشكل q^{x-1} . حيث $x = 1, 2, 3, \ldots$ ؛ سنرمز لهذا التوزيع بالرمز $g(x; p)$ أى أن احتمال الحصول على صورة فى القذفة الرابعة سيكون مساويا لـ :

$$g (4; \tfrac{1}{2}) = \left(\frac{1}{2} \right) . (\frac{1}{2})^{4-1} = \frac{1}{16}$$

مثال (٣,١٦)

تحتوى بضاعة مصنعة بالمتوسط بين كل مئة عنصر ثلاثة عناصر معابة . ما هو احتمال ظهور أول عنصر معيب عند فحص عاشر عنصر من عناصر هذه البضاعة ؟

الحل

باستخدام التوزيع السابق ، وبفرض أن $p = 0.03, x = 10$ فإننا نجد :

$$g(10; 0.03) = (0.03)(0.97)^9$$

$$= 0.0228$$

تمارين محلولة

تمرين (١)

ألقينا حجر نرد عشر مرات متتالية . لنعتبر أن النجاح فى كل رمية يوافق ظهور أحد العددين 3 أو 5 . ما هو احتمال الحصول على 3 أو 5 خمس مرات ، ثم ما هو توقع وتباين مرات النجاح فى هذه التجربة .

الحل

لنرمز لعدد مرات النجاح بالرمز X . نجد أن هذا المتغير حدانى ، كما نلاحظ أن عدد الاختبارات المجراة $n = 10$ ، واحتمال النجاح فى كل اختبار هو $P\{3, 5\} = \frac{2}{6}$ لذلك فالاحتمال المطلوب هو :

$$P[X = 5] = \binom{10}{5} \left(\frac{1}{3}\right)^5 \left(\frac{2}{3}\right)^5$$

$$= 0.13656$$

أما توقع وتباين عدد مرات النجاح X فيحسبان بوساطة العلاقتين :

$$\mu = n.p, \sigma^2 = n.p.q$$

أى أن :

$$\mu = 10 \cdot \frac{1}{3} = \frac{10}{3}$$

$$\sigma^2 = \frac{10}{3} \cdot \frac{2}{3} = \frac{20}{9}$$

تمرين (٢)

أسرة مؤلفة من أب وأم وستة أطفال . إذا علمت أن احتمال أن يكون طفلا ما فى الأسرة ذكرا هو $\frac{1}{2}$ ، فما هو احتمال أن يكون فى الأسرة ثلاثة ذكور ، وثلاث إناث ؟

الحل

يمكن النظر إلى الأطفال الستة فى الأسرة على أساس ست اختبارات ، ويوافق الذكر نجاح والأنثى فشل . لذلك فإن احتمال أن يكون عدد النجاحات ثلاثة يحسب بالعلاقة :

$$P[X = x] = f(x) = \binom{n}{x} p^x \cdot q^{n-x}$$

وبالتعويض عن $n = 6$, $p = \frac{1}{2}$, $q = \frac{1}{2}$, $x = 3$ فإننا نجد أن :

$$P[X = 3] = \binom{6}{3} \left(\frac{1}{2}\right)^3 \cdot \left(\frac{1}{2}\right)^3$$

$$= \frac{6!}{3!\,3!} \cdot \frac{1}{2^6}$$

$$= 0.3125$$

تمرين (٣)

يرمى أحمد على هدف بعيارات نارية . إذا علمت أن احتمال إصابته للهدف هو 0.7 فما هو احتمال :

(١) أن يصيب الهدف مرة واحدة إذا أطلق عليه سبع طلقات متتالية ؟

(٢) أن يصيب الهدف مرتين على الأقل عندما يطلق سبع طلقات متتالية ؟

الحل

لنرمز بـ X لعدد المرات التى أصاب بها الهدف . نجد أن للمتغير X توزيعا حدانياً معيناً بالعلاقة :

$$P[X = k] = \binom{7}{k} (0.7)^k \cdot (0.3)^{7-k}$$

(١) أما احتمال إصابته للهدف مرة واحدة عند إطلاق سبع طلقات فهو يساوى :

$$P[X = 1] = \binom{7}{1} (0.7) \cdot (0.3)^6 = 0.0035721$$

(٢) لايجاد احتمال أن يصيب الهدف مرتين على الأقل ، علينا أن نبحث عن مجموع الاحتمالات من أجل $k = 2, 3, 4, 5, 6$ ، أو أن نحسب مجموع احتمالات أن يصيب مرة وإلا يصيب أى مرة ، ونطرح المجموع من واحد .

نلاحظ أن احتمال عدم الإصابة هو :

$$P[X = 0] = \binom{7}{0} (0.7)^0 (0.3)^7$$

$$= (0.3)^7$$

$$= 0.0002187$$

وأخيرا فإن احتمال إصابته مرتين على الأقل هو :

$$P[X > 2] = 1 - P[X < 2]$$

$$= 1 - [P(X = 0) + P(X = 1)]$$

$$= 1 - 0.0035721 - 0.0002187$$

$$= 0.9962092$$

تمرين (٤)

أسرة مؤلفة من عشرة أطفال . إذا علمت أن احتمال وجود ذكر يساوى احتمال وجود أنثى

(١) ما هو التوقع الرياضى لعدد الذكور فى هذه الأسرة ؟

(٢) احسب احتمال وجود خمسة ذكور .

الحل

نلاحظ من الفرض أن $p = \frac{1}{2}$ ، وأن $n = 10$ ، وتوزيع الاحتمال لعدد الذكور فى الأسرة هو توزيع حدانى فيه $p = \frac{1}{2}$ و $n = 10$ إذا فالتوقع الرياضى (لعدد الذكور X) هو :

$$E(X) = n.p = 10. \frac{1}{2} = 5$$

أما احتمال وجود خمسة ذكور فيحسب من العلاقة التالية :

$$P[X = 5] = b(5; 10, \tfrac{1}{2}) = \binom{10}{5} \quad (\tfrac{1}{2})^5 \quad (\tfrac{1}{2})^5$$

$$= \frac{10!}{5!\,5!} \cdot (\frac{1}{2^{10}})$$

$$= 0.24609$$

تمرين (٥)

بفرض أن نسبة اللمبات المعيبة فى إنتاج مصنع للأنابيب الألكترونية هى 2% ، وإذا علمت أن هذا المصنع قد أنتج 5000 لمبة ، فما هو التوقع الرياضى والانحراف المعيارى لعدد اللمبات المعيبة ؟

الحل

لنرمز بـ Y لعدد اللمبات المعيبة الموجودة بين 5000 لمبة منتجة . نلاحظ أن للمتغير Y توزيعا حدانيا باحتمال نجاح قدره $p = \frac{1}{50}$ ، كما أن $n = 5000$ لذلك حسب قوانين التوقع والتباين نجد أن :

$$\mu = E(X) = n.p = (5000) . (\frac{1}{50}) = 100$$

$$\sigma^2 = E(X - \mu)^2 = n.p.q = 100 . (\frac{98}{100}) = 98$$

والانحراف المعيارى كما نعلم ما هو إلا الجذر الموجب للتباين ، ولذلك يكون :

$$\sigma = 9.899$$

تمرين (٦)

متغير عشوائى حدانى X توقعه 100 وانحرافه المعيارى 9.899 . ما هو توزيع هذا المتغير ؟

الحل

نعلم أن توزيع الاحتمال للمتغير الحدانى X يعطى بالعلاقة :

$$f(x) = \binom{n}{x} p^x . q^{n-x}$$

ولابد من تحديد n وتحديد p . لذلك نعلم أنه بالنسبة للمتغير الحدانى :

$$\mu = n.p \quad , \quad \sigma = \sqrt{np . (1-p)}$$

وحسب المعطيات نجد أن :

$$100 = n.p \quad , \quad 9.899 = \sqrt{np . (1-p)}$$

وهما معادلتين بمجهولين بحلهما نجد أن $p = \frac{2}{100}$. $n = 5000$ ، وبالتعويض فى عبارة f(x) ، فإننا نجد توزيع الاحتمال للمتغير المفروض X هو :

$$f(x) = \binom{5000}{x} \left(\frac{2}{100}\right)^x \cdot \left(\frac{98}{100}\right)^{5000-x}$$

والملاحظ فى هذه المسألة أن عبارة الكثافة الاحتمالية السابقة تمثل توزيع الاحتمال لعدد اللمبات المعيبة والموجودة ضمن 5000 لمبة إذا علمنا أن نسبة العيب هو $\frac{2}{100}$ بين إنتاج المصنع من اللمبات .

تمرين (٧)

يحوى صندوق 12 كرة حمراء ، 6 صفراء ، و 8 سوداء . سحبنا وبصورة عشوائية مع الإعادة خمس كرات من الصندوق ، ما هو احتمال أن يكون بين الكرات المسحوبة كرتان حمراء ، كرة صفراء ، وكرتان سوداوات ؟

الحل

نلاحظ أن الاحتمال المطلوب يمكن إيجاده بوساطة العلاقة :

$$f(x_1, x_2, x_3 ; p_1, p_2, x_3, n) = \frac{n!}{x_1! \, x_2! \, x_3!} P_1^{x_1} \cdot P_2^{x_2} \cdot P_3^{x_3}$$

كما نلاحظ أن :

$$p_1 = \frac{12}{26} \quad p_2 = \frac{6}{26} \quad p_3 = \frac{8}{26} \quad x_1 = 2 \quad x_2 = 1 \quad x_3 = 2 \quad n = 5$$

وبالتعويض نجد أن :

$$f\left(2, 1, 2; \frac{12}{26}, \frac{6}{26}, \frac{8}{26}, 5\right) = \frac{5!}{(2!)(1!)(2!)} \left(\frac{6}{13}\right)^2 \cdot \left(\frac{3}{13}\right)^1 \cdot \left(\frac{4}{13}\right)^2$$

$$= 0.13962$$

تمرين (٨)

قذفنا حجر نرد ست مرات والمطلوب :

(١) ما هو احتمال ظهور الواحد مرة ، الثلاثة مرتين ، والستة ثلاث مرات ؟

(٢) ما هو احتمال ظهور كل وجه مرة واحدة ؟

الحل

لدينا فى هذه المسألة تجربة متعددة الحدود .

(١) الاحتمال المطلوب في (١) يكتب بالشكل التالى :

$$f\left(1, 2, 3; \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, 6\right) = \frac{6!}{1!\,2!\,3!}\left(\frac{1}{6}\right)^1 \cdot \left(\frac{1}{6}\right)^2 \cdot \left(\frac{1}{6}\right)^3$$

$$= 0.001286$$

(٢) أما احتمال ظهور كل وجه مرة واحدة فيساوى :

$$f\left(1, 1, 1, 1, 1, 1; \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, 6\right) = \frac{5}{324} = 0.01543$$

تمرين (٩)

مجموعة بضاعة مصنعة مؤلفة من مئة عنصر تحتوى على عشرة عناصر معيبة . سحبنا وبصورة عشوائية عينة مؤلفة من خمس عناصر لدراسة صفتها النوعية . ما هو توزيع الاحتمال لعدد العناصر المعيبة والموجودة ضمن العينة المستخرجة ؟

الحل

نرمز لعدد العناصر المعيبة بـ X ، نلاحظ أن القيم التى يفترضها المتغير X هى :

$$0, 1, 2, 3, 4, 5$$

والملاحظ أن هذا المتغير له توزيع هندسى زائدى باحتمال قدره :

$$P[X = K] = \frac{\binom{10}{K} \cdot \binom{90}{5-K}}{\binom{100}{5}} : K = 0, 1, 2, 3, 4, 5$$

وهكذا نجد أن للمتغير X توزيعا محددا بالجدول التالى :

X	0	1	2	3	4	5
P	0.583	0.34	0.07	0.007	0	0

نلاحظ من الجدول السابق أن :

$$\sum_{i=0}^{5} P_i = 1$$

تمرين (١٠)

جملة منتجات مصنعة تخضع للفحص للنظر فى إمكانية صلاحيتها للتسويق خارج البلاد ، فإذا علمت أن احتمال اجتياز كل عنصر من هذه المنتجات الفحص (الاختبار) هو $\frac{4}{5}$ ، وأن هذه الاختبارات (الفحوص) مستقلة ، وأن عملية الفحص تقف عند ظهور أول عنصر لم يجتز الفحص . فما هو الاحتمال لعدد التجارب المجراة ؟

الحل

لنرمز بـ Y لعدد التجارب المجراة حتى ظهور أول عنصر لم يجتز الفحص ، نلاحظ أن للمتغير Y توزيعاً حدانياً سالباً خاصاً وذلك من أجل $k = 1$ أى أن :

$$P [Y = y] = g(y ; p) = (\frac{4}{5})^{y-1} . (\frac{1}{5}) : y = 1,2,3, \ldots$$

وهكذا نجد أن :

X	1	2	3	.... y ....
P	$\frac{1}{2}$	$\frac{4}{(5)^2}$	$\frac{(4)^2}{(5)^3}$	$\frac{(4)^{y-1}}{(5)^y}$

تمرين (١١)

إذا علمت أن احتمال معاناة مريض من رد فعل سيء عند حقنه بمصل معين هو 0.0001 ، فما هو احتمال :

(١) أن يعانى خمسة بالضبط من رد الفعل السيء من هذا الحقن .
(٢) أن يعانى أكثر من خمسة من رد الفعل السيء من هذا الحقن .

وذلك عند حقن 3000 مريض بهذا المصل .

الحل

إذا فرضنا أن Y يمثل عدد المرضى الذين عانوا من رد الفعل السيء من الحقن

بالمصل المذكور ، عندئذ نلاحظ أن لـ Y توزيعاً بواسونياً ، ولهذا فالاحتمال يحسب بالعلاقة :

$$P[Y = y] = e^{-\mu} \cdot \frac{\mu^y}{y!}$$

وبالنسبة لاحتمال معاناة خمسة مرضى من رد الفعل السيء فهو يساوى :

$$P[Y = 5] = e^{-\mu} \cdot \frac{\mu^5}{5!} : \mu = n \cdot p = (3000) \left(\frac{1}{10000}\right)$$

ومنه :

$$P[Y = 5] = e^{-\frac{3}{10}} \cdot \frac{\left(\frac{3}{10}\right)^5}{5!} = \frac{1.5}{(10)^5}$$

كذلك فإن احتمال أن يعانى أكثر من خمسة مرضى من رد الفعل السيء هو :

$$P[Y > 5] = 1 - P[Y \leq 5]$$
$$= 1 - \sum_{y=0}^{5} P\left(y ; \frac{3}{10}\right)$$

نلاحظ من الجدول III أن :

$$\sum_{y=0}^{y} P(y; 0.3) = 1$$

إذا

$$\sum_{y=0}^{5} p(y, 0.3) = 1$$

أيضا ، والاحتمال المطلوب معدوم .

تمرين (١٢)

تبلغ نسبة الكؤوس التالفة فى إنتاج مصنع زجاج خمسة فى المائة من إنتاجه ، ما هو احتمال وجود كأسين تالفين فى عينة مؤلفة من 12 كأسا مأخوذة بشكل عشوائى من جملة الإنتاج .

الحل

نلاحظ أن احتمال ظهور كأس تالف فى إنتاج المصنع هو $P = \frac{5}{100}$ ، فإذا أخذنا عينة من إنتاج المصنع حجمها $n = 12$ ، وبفرض أن X يمثل عدد القطع التالفة ضمن هذه العينة ، فإننا نجد أن لـ X توزيعا بواسونيا نظرا لكبر n وصغر p . ولذلك فإن الاحتمال المطلوب هو :

$$P[X = 2] = e^{-\mu} \cdot \frac{\mu^2}{2!} : \mu = n \cdot p = 12 \cdot \frac{5}{100} = 0.6$$

$$= e^{-0.6} \cdot \frac{(0.6)^2}{2!}$$

$$= 0.098786$$

تمرين (١٣)

تحتوى مساحة صغيرة من زجاجة مجهرية لفحص الدم من أجل أى شخص طبيعى فى المتوسط على عشرة كريات حمراء . ما هو احتمال أن تحتوى زجاجة من شخص طبيعى فى تلك المساحة الصغيرة على أقل من 6 كريات حمراء ؟

الحل

لنرمز بـ Y لعدد الكريات الحمراء التى تحويها المساحة الصغيرة من زجاجة مجهرية لفحص الدم من أجل شخص طبيعى فيكون المطلوب حساب :

$$P[Y < 6] = \sum_{x=0}^{5} P(y; \mu)$$

ولكن

$$E(Y) = \mu = 10$$

إذن :

$$P[Y < 6] = \sum_{x=0}^{5} P(y; 10)$$

ومن الجدول III نجد أن :

$$= 0.0671$$

تمارين عامة

(١) أوجد معادلة توزيع الاحتمال للمتغير العشوائى X الممثل لرقم الكرت المسحوب بشكل عشوائى من صندوق يحوى عشرة كروت مرقمة من الواحد إلى عشرة . ما هو احتمال أن يكون رقم الكرت المسحوب أقل من أربعة ؟

(٢) أوجد التوقع الرياضى والتباين للمتغير العشوائى X فى التمرين .

(٣) قسمت عجلة الروليت إلى 25 قطاعان بمساحات متساوية من الواحد إلى خمس وعشرين . أوجد عبارة توزيع الاحتمالى للمتغير X الممثل للعدد الذى ظهر عند إدارة عجلة الروليت .

(٤) إذا علمت أن 75% من السيارات التى عبرت نقطة سير ضوئية كانت من داخل مدينة جدة فى المملكة العربية السعودية . ما هو احتمال أن تكون ثلاث على الأقل من (خمسة سيارات أخرى ستعبر الإشارة الضوئية) من خارج مدينة جدة ؟

(٥) إذا علمت أن 75% من دجاج مدجنة معينة ملقحة بلقاح ضد مرض يصيب الدجاج . فإذا لقحنا ثلاث دجاجات ، فما هو احتمال أن تكون دجاجتان على الأكثر قد التقطتا المرض ؟

(٦) سحبنا ورقة من ورق اللعب المؤلف من 52 ورقة ثم أعدناها وكررنا التجربة خمس مرات متتالية . ما هو احتمال أن نحصل على ورقتين دينارى وواحدة بستونى ؟

(٧) احسب احتمال الحصول على الأرقام 1, 2, 3, 4, 5, 6 عددا من المرات 2, 1, 1, 2, 3, 1 على الترتيب ، وذلك عند إلقاء حجر نرد عشر مرات متتالية ؟

(٨) سحبنا أربعة صواريخ من مجموعة مؤلفة من عشرة صواريخ وذلك بشكل عشوائى ، وقذفناها بوساطة مدفع . فإذا علمت أن مجموعة العشرة صواريخ تحتوى على ثلاثة صواريخ غير قابلة للانفجار فما هو احتمال :

أ – أن تنفجر الصواريخ الأربعة المختارة ؟

ب – اثنان على الأكثر لم ينفجرا ؟

(٩) فى التمرين الثامن . ما هو عدد الصواريخ المعيبة التى تتوقع ألا تنفجر بين الأربعة المسحوبة بشكل عشوائى . استخدم متباينة تشبيشيف لوصف تغيرات هذا العدد ؟

(١٠) من المقدر أن يكون 4000 صوت من أصوات الناخبين البالغ عددهم 10000 ناخب ضد مشروع فرض ضريبة على الأرباح . فإذا اخترنا 15 شخصا ممن يحق لهم التصويت وتم سؤالهم عن رأيهم ، فما هو احتمال أن يكون سبعة على الأكثر مع مشروع فرض الضريبة المذكورة ؟

(١١) إذا علمت أن متوسط عدد حوادث السير خلال أسبوع عند إشارة ضوئية معينة هو ثلاثة ، فما هو احتمال أن تقع خمسة حوادث بالضبط عند هذه الإشارة خلال الأسبوع القادم ؟

(١٢) تم تشكيل اتحاد للطلاب المسلمين من أربعة أقطار عربية ، وقد انضم إلى هذا الاتحاد ثلاثة طلاب سعوديين ، خمسة فلسطينيين ، وطالبين سوريين وطالبين مصريين . ما هو احتمال أن تمثل الجنسيات الأربعة فى لجنة مؤلفة من أربعة طلاب مختارة بشكل عشوائى ؟

(١٣) إذا علمت أن متوسط عدد الأخطاء التى ترتكبها سكرتيرة عند طباعتها الورقة على الآلة الكاتبة هو اثنتان ، فما هو احتمال :

أ – أن ترتكب هذه السكرتيرة أربعة أخطاء أو أكثر فى الصفحة الثانية .

ب – ألا ترتكب أى خطأ فى الصفحة الثانية .

(١٤) إذا علمت أن احتمال موت شخص فى عدوى تصيب الجهاز التنفسى هو 0.002 . أوجد احتمال أن يكون أكثر من خمسة أشخاص من أصل 2000 شخص مصابين بهذه العدوى قد ماتو .

(١٥) ما هو احتمال الحصول على صورة ثالثة فى الالقاءة السابعة لدى إلقاء قطعة نقود سبع مرات متتالية ؟

(١٦) للحصول على رخصة قيادة يتقدم المواطن للفحص أمام لجنة ، فقد يجتاز

هذا الفحص وقد يفشل . إذا علمت أن احتمال أن اجتياز مواطن الفحص فى كل مرة هو 0.7 ، فما هو احتمال أن يجتاز سعيد فحص القيادة هذا فى :

أ – المرة الثالثة .

ب – قبل المرة الرابعة .

الفصل الرابع

بعض توزيعات الاحتمال المستمرة

■ التوزيع الطبيعي ■ المساحة تحت المنحنى الطبيعي ■ التقريب الطبيعي للتوزيع الحدانى ■ التوزيعات غما ، الأسي ، كاى-مربع ■ توزيع وايبل ■ تمارين محلولة ■ تمارين عامة .



(٤,١) التوزيع الطبيعى Normal distribution

تمهيد

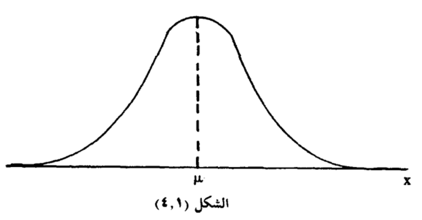
كما ذكرنا سابقا فإن المتغير العشوائى المستمر يمسح جميع نقاط محور موجه ، وبالتالى فإنه بالإضافة لكون عدد قيم هذا المتغير لانهائياً ، فإن هذا العدد غير قابل للعد كنقاط المجال (a, b) مثلا . وكبعض الأمثلة على المتغيرات المستمرة نذكر على سبيل المثال عمر مصباح كهربائى ، أخطاء القياسات فى تجربة مخبرية ، طول إنسان ... إلخ للحصول على نموذج احتمالى لمتغير مستمر نبدأ باختبار منحن مستمر يمثل ما يسمى بدالة الكثافة الاحتمالية f(x) . ولا بد لهذه الدالة أن تحقق الشرطين التاليين :

١) $f(x) \geq 0$ من أجل جميع قيم x

٢) المساحة بين المنحنى f(x) والمحور السينى تساوى الواحد .

وعندئذ سيكون احتمال أى حادث ممثلا للمساحة تحت منحنى الكثافة . ونتيجة لذلك ، فإن احتمال فرض أن المتغير X مساوياً للقيمة a مثلا $P(X = a)$ هى المساحة تحت المنحنى فوق النقطة a من المحور السينى وهذه المساحة صفر . وهكذا نجد أن احتمال أن يأخذ المتغير العشوائى المستمر (قيمة بحد ذاتها) معدوم . وهذا تعبير واقعى عن استحالة توصل الإنسان إلى أجهزة قياس دقيقة بصورة مطلقة . لذلك تبقى مثل هذه النتيجة مقبولة طالما بقى الإنسان غير قادر على الادعاء بأن قياساته لعمر مصباح كهربائى ، مثلا ، هى قياسات لا تخضع لأى خطأ على الإطلاق ، إذ مهما أوتى جهاز القياس من الدقة ، ومهما بلغت مهارة الإنسان الذى يستخدم هذا الجهاز فلابد من ارتكاب خطأ مهما كان صغيراً .

وبينما تتخذ منحنيات الكثافة أشكالا مختلفة ، نلاحظ أن عدداً كبيراً من المتغيرات العشوائية التى تصادفنا فى تجاربنا اليومية لها منحنى كثافة له تقريبا شكل الجرس ، أو كما نعبر عن ذلك إحصائياً ، له بصورة تقريبية ، شكل التوزيع الطبيعى كما هو موضح فى الشكل (٤,١) .



الشكل (٤,١)

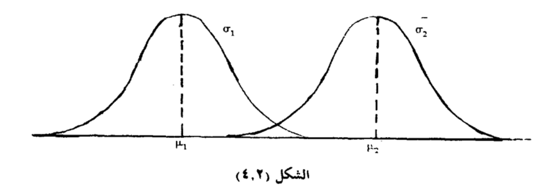
يدعى المتغير العشوائى ذو التوزيع الجرسى كما فى الشكل (٤,١) بالمتغير الطبيعى ، والمعادلة الرياضية لهذا التوزيع تتعلق بمتحولين σ, μ (أى المتوسط والانحراف المعيارى للمتغير الطبيعى ، لذلك سنرمز للكثافة الاحتمالية لهذا المتغير بالرمز $n\ (x; \mu, \sigma)$

التوزيع الطبيعى

إن الكثافة الاحتمالية للمتغير العشوائى الطبيعى X ذى التوقع μ ، والانحراف المعيارى σ تعطى بالعلاقة :

$$n\ (x; \mu, \sigma) = \frac{1}{\sqrt{2\pi} \cdot \sigma}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad : -\infty < x < +\infty$$

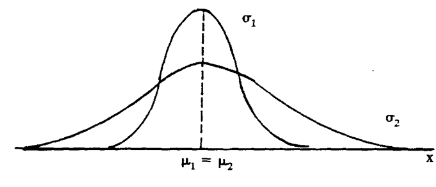
حيث يمثل 3.14159 = $e = 2.71828$. يكون المنحنى الطبيعى محدداً تماماً إذا كانت σ, μ محددتين . مثلا إذا كان $\sigma = 2, \mu = 36$ عندئذ يمكن بسهولة حساب $n(x; 36, 2)$ لجميع قيم x ويمكن رسم هذا المنحنى . يوضح الشكل (٤,٢) منحنيين طبيعيين لهما نفس الانحراف المعيارى σ بتوقعين مختلفين .



الشكل (٤,٢)

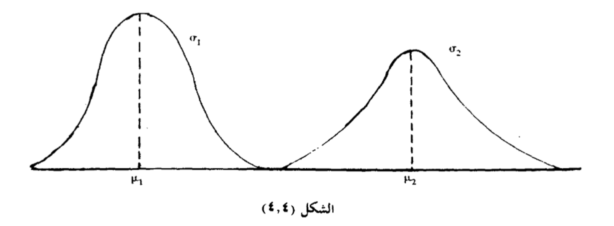
ونلاحظ أن المنحنيين متطابقان بالشكل ، ولكنهما متمركزين فى نقطتين مختلفتين على طول محور الفواصل .

يوضح الشكل (٤,٣) منحنيين طبيعيين لهما نفس التوقع (الوسط) ، وانحرافين مختلفين ، كما نلاحظ على نفس الشكل أن المنحنيين متمركزان فى نفس نقطة على محور الفواصل ، ولكن المنحنى ذو الانحراف المعيارى الأكبر أخفض من المنحنى الآخر .



الشكل (٤,٣)

أما الشكل (٤,٤) فيوضح الرسم البيانى لمنحنيين طبيعيين لهما توقعين مختلفين وانحرافين مختلفين أيضا ، ومن الواضح أنهما متمركزان فى نقطتين مختلفتين على محور الفواصل وشكليهما يعكسان اختلاف انحرافيهما المعياريين .



الشكل (٤,٤)

$$\mu_1 > \mu_2 \, , \; \sigma_1 < \sigma_2$$

بالعودة إلى المشتقتين الأولى والثانية للدالة $n(x; \mu, \sigma)$ مقارنة الأشكال السابقة نستنتج الخواص التالية التي يتمتع بها المنحنى الطبيعي .

(١) يبلغ المنحنى نهايته العظمى فى النقطة $x = \mu$ من المحور السينى .

(٢) المنحنى الطبيعي متناظر حول محور التراتيب المار من النقطة μ .

(٣) يعانى المنحنى الطبيعي انعطافا فى النقاط $x = \mu \pm \sigma$ ، كما أن تقعره يكون نحو الأسفل إذا كان $\mu - \sigma < X < \mu + \sigma$. ونحو الأعلى فيما عدا ذلك .

(٤) يصبح المنحنى الطبيعي متقاربا من محور الفواصل كلما ابتعدنا عن μ فى مختلف الاتجاهات .

(٥) المساحة تحت المنحنى الطبيعي وفوق محور الفواصل تساوى الواحد .

سنبرهن أن μ , σ^2 يمثلان على الترتيب توقع وتباين المتغير العشوائى الطبيعي . ولحساب التوقع الرياضى نكتب الآتي :

$$E(X) = \frac{1}{\sqrt{2\pi}\,.\,\sigma}\int_{-\infty}^{+\infty} x\,.\,e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}dx$$

وباجراء تغيير للمتحول بالعلاقة $Z = \frac{x-\mu}{\sigma}$ نجد أن $\sigma . dz = dx$ و $X = \sigma . Z + \mu$ وبالتعويض نجد أن :

$$E(X) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}(\sigma\,.\,Z+\mu)\,.\,e^{-\frac{1}{2}Z^2}\,dz$$

$$= \frac{\mu}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{-\frac{1}{2}Z^2}\,dz + \frac{\sigma}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} Z\,e^{-\frac{1}{2}Z^2}\,dz$$

ومن الواضح أن التكامل $\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{-\frac{1}{2}Z^2}\,dz$ يمثل المساحة الواقعة تحت المنحنى الطبيعي ذى التوقع صفر والانحراف المعيارى واحد لذلك فقيمته الواحد . ومنه $\mu . 1 = \mu$ ، أما التكامل الثانى فيبرهن أنه يساوى الصفر . وهكذا نجد أن :

$$E(X) = \mu . 1 + 0 = \mu$$

أما بالنسبة للتباين ، فمن المعلوم أن :

$$E[(X-\mu)^2] = \frac{1}{\sqrt{2\pi}\cdot\sigma}\int_{-\infty}^{+\infty}(x-\mu)^2 \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

وبتغير المتحول بوساطة العلاقة السابقة نفسها نجد أن :

$$= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}\sigma^2 \cdot Z^2 e^{-\frac{1}{2}Z^2} dz$$

$$= \frac{\sigma^2}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} Z^2 \cdot e^{-\frac{1}{2}Z^2} dz$$

وبالتكامل بالتجزئة على اعتبار أن $U = Z$ و $dv = Z \cdot e^{-\frac{1}{2}Z^2} dZ$ نجد أن :

$$E[(X-\mu)^2] = \frac{\sigma^2}{\sqrt{2\pi}}\left[-Z \cdot e^{-\frac{Z^2}{2}}\Big|_{z=-\infty}^{+\infty} + \int_{-\infty}^{+\infty} e^{-\frac{1}{2}Z^2} dz\right]$$

$$= \frac{\sigma^2}{\sqrt{2\pi}}[0 + \sqrt{2\pi}]$$

$$= \sigma^2$$

(٤,٢) المساحة تحت المنحنى الطبيعي

Area under the normal distribution

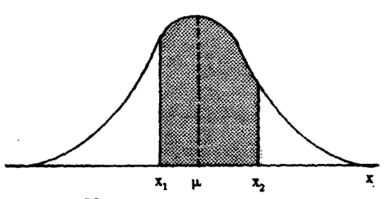
من المعلوم أن المساحة تحت منحنى التوزيع الاحتمالي المستمر ، وبين أي نقطتين $x = x_1$ $x = x_2$ يمثل احتمال أن يتردد المتغير المفروض بين هاتين النقطتين أي أن :

$$P[x_1 < X < x_2] = \int_{x_1}^{x_2} f(x)\, dx$$

ومن أجل التوزيع الطبيعي نجد أن :

$$P[x_1 < X < x_2] = \frac{1}{\sqrt{2\pi}\cdot\sigma}\int_{x_1}^{x_2} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \cdot dx$$

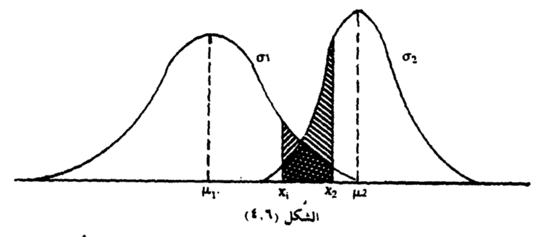
وهذا التكامل يمثل مساحة القسم المظلل من الشكل (٤,٥) .



$P[x_1 < X < x_2]$ = مساحة القسم المظلل

الشكل (٤,٥)

لقد رأينا من خلال الأشكال (٤,٢) ، (٤,٣) ، (٤,٤) كيف أن المنحنى الطبيعى يعتمد على التوقع والانحراف المعيارى للتوزيع ، وأن المساحة تحت المنحنى بين أى نقطتين تعتمد أيضا قيم كل من μ ، σ ، وهذا ما يوضحه الشكل (٤,٦) ، حيث ظللنا المنطقة المقابلة لـ $P[x_1 < X < x_2]$ حيث يمثل X متغيرا عشوائيا يصفه التوزيع I يمثله القسم المظلل من نفس الشكل . أما إذا كان X متغيراً عشوائياً موصوفاً بالتوزيع II ، فإن هذا الاحتمال يمثله القسم المظلل من نفس الشكل II . وبما أن المساحتين المظللتين على كل شكل مختلفتان ، لذلك فإن الاحتمال سيختلف تبعا لشكل توزيع المتغير X .



الشكل (٤,٦)

ولحسن الحظ أنه يمكننا أن نغير كافة الملاحظات المتعلقة بمتغير عشوائى ما X إلى مجموعة ملاحظات متغير عشوائى آخر Z ذو التوقع صفر والتباين واحد . ويمكن إجراء ذلك بواسطة علاقة التحويل :

$$Z = \frac{X - \mu}{\sigma}$$

فإذا افترض المتغير X القيمة x ، فعندئذ يفترض المتغير Z القيمة $\frac{x-\mu}{\sigma}$.
إذا وقع X بين القيمتين x_1, x_2 فإن المتغير الجديد Z سيقع بين القيمتين $\frac{x_2-\mu}{\sigma}$ و $\frac{x_1-\mu}{\sigma}$ على الترتيب . ولهذا يمكن أن نكتب :

$$P[x_1 < X < x_2] = \frac{1}{\sqrt{2\pi}\,\sigma}\int_{x_1}^{x_2} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} . dx$$

$$= \frac{1}{\sqrt{2\pi}}\int_{z_1}^{z_2} e^{-\frac{1}{2}z^2}\, dz$$

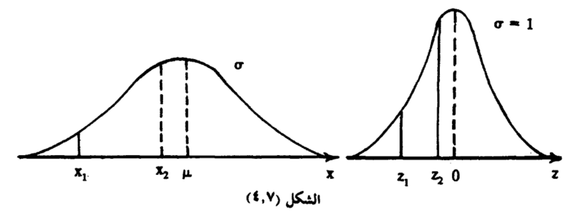
$$= \int_{z_1}^{z_2} n(z; 0, 1)\, dz = P[z_1 < Z < z_2]$$

حيث رمزنا بـ z للمتغير الطبيعى ذى التوقع الصفرى والتباين المساوى للواحد .

تعريف (٤,١) التوزيع الاحتمالى Probability distribution

يدعى التوزيع الاحتمالى للمتغير العشوائى الطبيعى بالتوقع صفر والتباين واحد بالتوزيع الطبيعى المعيارى .

يوضح الشكل (٤,٧) كلا من التوزعين الطبيعى والطبيعى المعيارى .



الشكل (٤,٧)

يعطينا الجدول IV المساحة الواقعة أسفل التوزيع الطبيعى المعيارى الموافقة للاحتمال $P[Z < 2]$. وذلك من أجل جميع قيم Z من 3.4 - إلى 3.4 . سنذكر فيما يلى طرق استخدام هذا الجدول .

مثال (٤,١)

X متغير عشوائى طبيعى بالتوقع $\mu = 36$ ، والتباين $\sigma^2 = 4$ ، لحساب $P[30 < X < 37]$ ، أى المساحة الواقعة تحت المنحنى الطبيعى المعيارى السابق بين النقطتين 37,30 ، فإننا نقوم بتحويل هذا التوقع إلى التوزيع الطبيعي بواسطة علاقة التحويل :

$$Z = \frac{x - \mu}{\sigma} = \frac{x - 36}{2}$$

نلاحظ أنه يقابل النقطتين $x_1 = 30, x_2 = 37$ النقطتان :

$$Z_1 = \frac{30 - 36}{2} = -3$$

$$Z_2 = \frac{37 - 36}{2} = 0.5$$

لذلك فإن :

$$P[30 < X < 37] = P[-3 < Z < 0.5]$$
$$= P[Z < 0.5] - P[Z < -3]$$

وباستخدام الجدول IV من أجل $Z = 0.5$ ، من أجل $Z = -3$ نجد أن :

$$= 0.3085 - 0.0013$$
$$= 0.2072$$

ملاحظة هامة

بفرض أن للمتغير العشوائى الطبيعى X توقعاً μ وتبايناً σ^2 فإننا نلاحظ أن :

$$P[\mu - 2\sigma < X < \mu + 2\sigma] = P\left[\frac{\mu - 2\sigma - \mu}{\sigma} < \frac{X - \mu}{\sigma} < \frac{\mu + 2\sigma - \mu}{\sigma}\right]$$

$$= P[-2 < Z < 2]$$

$$= P[Z < 2] - P[Z < -2]$$

ومن الجدول IV ، نجد أن : من أجل $Z = 2$ و $Z = -2$:

$$P[\mu - 2\sigma < X < \mu + 2\sigma] = 0.9772 - 0.0228$$

$$P[\mu - 2\sigma < X < \mu + 2\sigma] = 0.9544$$

والمتباينة الأخيرة أفضل وأقوى من متباينة تشبيشيف . نخلص للقول بأن شكل متباينة تشبيشيف من أجل المتغيرات الطبيعية بالتوقع μ والانحراف المعيارى σ تأخذ الشكل الجديد التالى :

$$P[\mu - 2\sigma < X < \mu + 2\sigma] = 0.9544$$

مثال (٤,٢)

رُكِبَتْ فى إحدى القواعد الحربية مدفعية أرضية ، فإذا علمت أن عمر هذه المدفعية (أى صلاحيتها للعمل) يتوزع وفقا للتوزيع الطبيعى بالمتوسط 4 سنوات ، والانحراف المعيارى 0.6 سنة ، فاحسب احتمال أن تظل هذه المدفعية صالحة للعمل فى أقل من 3.3 سنوات .

الحل

لنرمز لفترة عمل المدفعية السابقة بالرمز X . نلاحظ أن $Z = \frac{X - 4}{0.6}$ يمثل متغيراً طبيعياً معيارياً والاحتمال المطلوب هو :

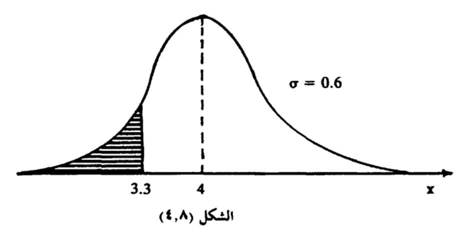
$$P[X < 3.3] = P\left[Z < \frac{3.3 - 4}{0.6}\right]$$

$$= P[Z < -1.17]$$

وبالعودة إلى الجدول IV ، نجد أنه يقابل القيمة $Z = -1.17$ مساحة قدرها 0.1210 تحت المنحنى الطبيعي المعياري ، ولذلك فالاحتمال المطلوب هو :

$$P[X < 3.3] = 0.1210$$

يوضح الشكل (٤,٨) هذه المساحة بالقسم المظلل من الشكل .



الشكل (٤,٨)

مثال (٤,٣)

بفرض أن لنوع معين من المصابيح الكهربائية عمراً يتوزع بشكل طبيعي بالوسط 750 ساعة ، والانحراف المعياري 60 ساعة . ما هو احتمال احتراق لمبة من هذه اللمبات بين 700 و 780 ساعة من اشتعالها ؟

الحل

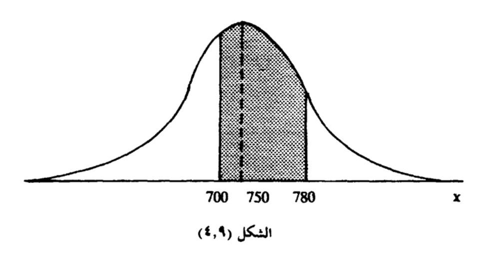
يوضح الشكل (٤,٨) توزيع عمر هذا النوع من المصابيح المنتجة ، ونلاحظ أن قيم Z الموافقة لـ $x_1 = 700, x_2 = 780$ هي :

$$Z_1 = \frac{700 - 750}{60} = -0.83$$

$$Z_2 = \frac{780 - 750}{60} = 0.5$$

لذلك فالاحتمال المطلوب هو :

$$P[700 < X < 780] = P[-0.83 < Z < 0.5]$$
$$= P[Z < 0.5] - P[Z < -0.83]$$
$$= 0.6915 - 0.2033$$
$$= 0.4882$$



الشكل (٤,٩)

مثال (٤,٤)

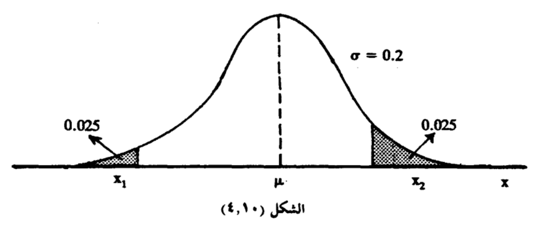
استخدمت مجموعة من القياسات لرفض جميع المركبات غير الموجودة فى المجال (1.5 - d, 1.5 + d) لبعد معين . فإذا عُلِم أن هذه القياسات توزيعاً طبيعياً بالمتوسط 1.5 والانحراف المعيارى 0.2 . فما هى قيمة d بحيث يغطى المجال المذكور 95% من هذه القياسات ؟

الحل

لنحدد قبل كل شيء قيمة كل من Z_2, Z_1 بحيث يكون :

$$p(Z_1 < Z < Z_2) = 0.95$$

نلاحظ من الشكل (٤,١١) ومن الجدول IV أن $Z_2 = 1.96$, $Z_1 = -1.96$



الشكل (٤,١٠)

لنحدد الآن قيمة $x_2 = \sigma . Z_2 + \mu$ بحيث تكون :

$$x_2 = 1.5 + d = (0.2)\,(1.96) + 1.5$$

من هذه العلاقة نجد أن $d = 0.392$

مثال (٤,٥)

تُنتج آلة معينة مقاومات كهربائية لها وسط مقاومة 35 أوم وانحراف معيارى قدره 1.8 أوم . بفرض أن المقاومة تتوزع طبيعياً ، ويمكن أن تقاس بالنسبة لأى درجة من الدقة ، فما هى النسبة المئوية للمقاومات التى سيكون لها مقاومة أكبر من 38 أوم ؟

الحل

نلاحظ أنه يمكن إيجاد النسبة المئوية بضرب التكرار النسبى بـ 100% . وبما أن التكرار النسبى من أجل أى مجال يساوى احتمال الوقوع فى هذا المجال ، لذلك يجب إيجاد المساحة الواقعة على يمين النقطة $x = 38$ كما هو موضح على الشكل (٤,١١) وهذا بالطبع يمكن تحديده بواسطة تحويل $x = 38$ إلى القيمة الموافقة $Z = \frac{38 - 35}{1.8}$

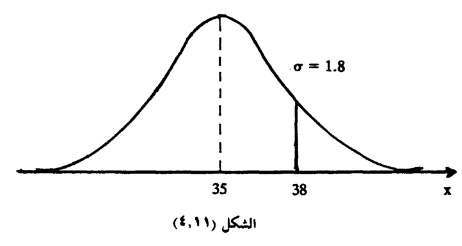
$$P[X > 38] = P[Z > 1.67]$$

$$= 1 - P\,[Z < 1.67]$$

ومن الجدول IV نجد ان :

$$= 1 - 0.9527$$
$$= 0.0473$$

لذلك فإن 4.7% من المقاومات سيكون لها مقاومة تفوق 38 أوم .



الشكل (٤,١١)

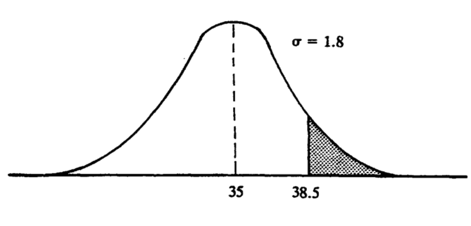
مثال (٤,٦)

ما هى النسبة المئوية للمقاومات التى تفوق 38 أوم فى المثال السابق إذا علمت أن المقاومة مقاسة بالنسبة لأقرب أوم ؟

الحل

جوهر هذا المثال يختلف عن المسألة الواردة فى المثال (٤,٥) ففى المثال السابق ربطنا بكل مقاومة قياسا قدره 38 أوم . بالنسبة لكل المقاومات التى لها قيمة أكبر من 37.8 وأقل 38.5 أوم . ونحن متأكدون من تقريب التوزيع المنقطع باعتباره توزيعا طبيعياً مستمراً . والمساحة المطلوبة هى المنطقة المظللة على يمين النقطة 38.5 من الشكل (٤,١٢) . ونجد أن :

$$Z = \frac{38.5 - 35}{1.8} = 1.944$$



الشكل (٤,١٢)

لذلك فإن :

$$\begin{aligned} P[X > 38.5] &= P[Z > 1.944] \\ &= 1 - P[Z < 1.944] \\ &= 1 - 0.9738 \\ &= 0.262 \end{aligned}$$

من هذا نستنتج أن 2.6% من المقاومات تفوق 38 أوم عندما تقاس بالنسبة لأقرب أوم . ويمثل الفرق 2.1% = 2.6% − 4.7% بين الإجابتين فى المثالين كل المقاومات التى لها قيم أكبر من 38 وأقل من 38.5 أوم والتى اعتبرناها تساوى 38 أوم .

(٤,٣) التقريب الطبيعى للتوزيع الحدانى
Normal Approximation to the Binomial

رأينا فى الفصل السابق تطبيقات متعددة للتوزيع الحدانى ، وحسبنا فيها جميعا احتمال أن يأخذ المتغير X (الممثل لعدد النجاحات فى n اختباراً) قيمة معينة ، وقد اقتصرنا فى دراستنا على قيم n الصغيرة ، بسبب المشتقة فى حساب $b(x; n, p)$ عندما تكون n كبيرة (وفى مثل هذه احالة علينا أن نحسب الاحتمال الحدانى بوساطة إجراءات التقريب) ودرسنا أيضا فى الفقرة (٣,٥) طريقة تقريب التوزيع الحدانى إلى التوزيع الطبيعى عندما تكون n كبيرة و p قريبة من الصفر أو الواحد ، ولاحظنا عندئذ أن كلا

من التوزيع البواسونى والحدانى هو توزيع منقطع ، وفى المثال (٤,٦) كنا قد بينا أول تطبيق حول تقريب التوزيع الحدانى إلى الطبيعى . سنذكر فيما يلى نظرية تدعى بنظرية النهايات المركزية تقرب لنا التوزيع الحدانى إلى التوزيع الطبيعى . وذلك فى الحالة التى تكون فيها n كبيرة بقدر كاف و p ليست قريبة من الصفر أو الواحد ، وسنذكر هذه النظرية بدون برهان .

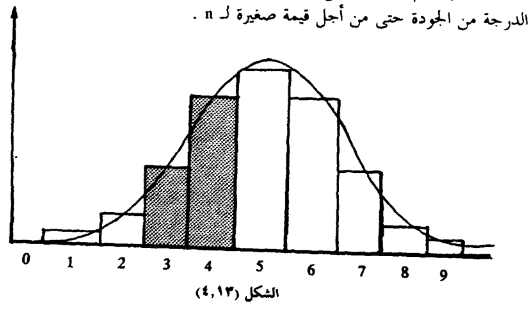
نظرية (٤,١)

إذا كان X متغيراً عشوائياً حدانياً بالتوقع $\mu = np$ ، والتباين $\sigma^2 = npq$ عندئذ يكون للمتغير :

$$Z = \frac{X - np}{\sqrt{np.q}}$$

توزيع يسعى إلى التوزع الطبيعى المعيارى وذلك عندما تسعى n إلى ∞ .

على سبيل المثال ، لنعتبر التوزيع الحدانى من أجل $n = 10$ و $p = \frac{1}{2}$. عندئذ نجد أن $\mu = np = 5$ و $\sigma = \sqrt{np.q} = 1.58$. يبين الشكل (٤,١٣) الاحتمال الموافق لحادث معين ، وذلك عند استخدام كل من التوزيع الحدانى والتوزيع الطبيعى ، كما تحدده نظرية النهايات المركزية . ونظرة أولية للشكل يتبين أن التقريب جيد تماما حتى فى الحالة $n = 10$. ويسهم تناظر التوزيع الحدانى فى حالة $n = 10$ فى كون التقريب على هذه الدرجة من الجودة حتى من أجل قيمة صغيرة لـ n .



الشكل (٤,١٣)

واحتمال أن يكون X مساوياً لـ 3 أو 4 يساوى تماما مساحة المستطيلين المقامين فوق $x = 3$ و $x = 4$. ويمكن تقريب هذه المساحة تحت المنحنى الطبيعى من $x = 2.5$ إلى $x = 4.5$ وهى المساحة المظللة فى الشكل (٤,١٣) .

والملاحظ أنه عندما تكون n صغيرة و p قريبة من الصفر أو الواحد ، فإن شكل المضلع الاحتمالى سيكون منحازاً (أى يتجمع معظمه) إلى جانب القيمة $x = 0$ أو $x = n$ على الترتيب . أى إنه سيكون بعيداً جداً عن وضع التناظر . وفى مثل هذه الحالات سيكون التقريب سيئا . وبصورة عامة كلما ابتعدت p عن القيمة $\frac{1}{2}$ كلما ابتعد شكل المضلع الاحتمالى للتوزيع الحدنى عن التناظر .

وإذا تذكرنا ما أشرنا إليه من أن 95% تقريبا من قياسات المجتمع الطبيعى ستقع ضمن المجال $(\mu \pm 2\sigma)$ ، فإننا نستنتج أن القياسات فى المجتمع الحدانى ستكون منتشرة فوق المجال $(\mu \pm 2\sigma)$. ولتحديد متى سيكون التقريب باستخدام التوزيع الطبيعى مناسباً أم لا ، نحسب $\mu = np$ و $\sigma = \sqrt{npq}$ فإذا وقع المجال $\mu \pm 2\sigma$ ضمن مدى التوزيع الحدانى ، أى بين 0 و n فسيكون التقريب جيداً .

مثال (٤,٧)

لنحسب فى التجربة الحدانية الموضحة على الشكل (٤,١٣) حيث فرضنا $n = 10$ ، $P = \frac{1}{2}$ احتمال أن يكون X مساويا 2 أو 3 أو 4 بدقة تصل الرقم العشرى الرابع ، وذلك باستخدام جدول التوزيع الحدانى II ، ثم باستخدام التقريب الطبيعى للتوزيع الثنائى .

الحل

نلاحظ أن :

$$P_1 = P[X = 2 \text{ أو } X = 3 \text{ أو } X = 4] = \sum_{x=0}^{4} b\left(x; 10, \frac{1}{2}\right) - \sum_{x=0}^{1} b\left(x; 10, \frac{1}{2}\right)$$

$$= 0.3770 - 0.0108$$

وهكذا نجد أن :

$$P_1 = 0.3662$$

أما إذا استخدمنا التقريب الطبيعي للتوزيع الحدانى ، فإننا نقوم بحساب المساحة بين $x_1 = 1.5$ ، $x_2 = 4.5$. وهكذا نكتب :

$$P_2 = P[1.5 < X < 4.5] = P\left[\frac{1.5 - 5}{1.58} < \frac{X - \mu}{\sigma} < \frac{4.5 - 5}{1.58}\right]$$

$$= P[-2.22 < Z < 0.32]$$

$$= P[Z < -0.32] - p[Z < -2.22]$$

$$= 0.4868 - 0.1255 = 0.3613$$

وبذلك نحصل على : $P_2 = 0.3613$

ونلاحظ أن القيمة P_2 تتفق مع القيمة الحقيقية P_1 برقمين عشريين .

مثال (٤,٨)

اخترنا لقاحاً ضد الزكام . وقد أعطى اللقاح لمئتى شخص ، وتمت مراقبتهم بالنسبة لإصابتهم بالزكام لمدة عام ، وقد نجا منهم 120 شخصاً من الإصابة . فإذا فرضنا أن احتمال عدم الإصابة بالزكام بصورة طبيعية دون استخدام أى لقاح هو 0.5 ، فما هى النتيجة التى يمكن استخلاصها من هذه التجربة حول فعالية اللقاح ؟

الحل

لنرمز بـ $p = 0.5, n = 200$. باستخدام التقريب الطبيعى للتوزيع الحدانى لنرمز بـ X لعدد الناجين من الإصابة بالزكام ، ولنحسب احتمال أن يكون X أكبر من أو يساوى 120 .

نلاحظ أن :

$$\sigma = \sqrt{np.q} = \sqrt{(200)\ (0.5)\ (0.5)} = 7.07$$

$$\mu = np = 200\ (0.5) = 100$$

$$P\ [X \geq 120] = p\left[\frac{X-\mu}{\sigma} \geq \frac{120-100}{7.07}\right]$$

$$= P\ [Z \geq 2.828]$$

$$= 1 - P\ [Z < 2.828]$$

$$= 1 - 0.9977$$

كما نلاحظ أيضا هذا الاحتمال من الصفر بحيث يمكن إهماله ، وهذا يدعونا إلى الاعتقاد بأن اللقاح مفيد بالفعل فى منع الإصابة بالزكام .

مثال (٤,٩)

تحتوى مجموعة بضاعة مصنعة على خمسة بالمئة من القطع المعابة . فإذا اخترنا وبصورة عشوائية مئة قطعة من هذه المجموعة ، فما هو احتمال أن تحوى هذه المجموعة الصغيرة أكثر من 6 عناصر معابة ؟

الحل

لنرمز بـ Y لعدد العناصر المعابة والموجودة فى العينة المستخرجة من البضاعة . نلاحظ أن للمتغير Y توزيعاً حدانياً بالمتغيرين $n = 100$ ، $P = 0.05$. بما أن حجم العينة n كبيراً ، فإننا نستخدم التقريب الطبيعى الحدانى حيث نأخذ :

$$\mu = n.p = 100\ .\ (0.05) = 5$$

$$\sigma = \sqrt{np.q} = \sqrt{(100)\ (0.05)\ (0.95)} = 2.179$$

ومنه :

$$P\,[Y > 6] = P\left[\frac{Y-\mu}{\sigma} > \frac{6-5}{2.179}\right] = P[Z > 0.458]$$

$$= 1 - P\,[Z < 0.458]$$

$$= 1 - 0.6772$$

$$= 0.3228$$

وقد يسأل سائل ، أما كان بالإمكان حل هذه المسألة بطريقة الجدول II . والجواب يبدو واضحاً إذا لاحظنا :

$$P\,[\,Y > 6] = \sum_{y=7}^{100} b\,(y;\,100,\,0.5)$$

إذ إن الجدول II صمم من أجل قيم n من الصفر وحتى $n = 20$. ومن المتعذر استخدام الجدول المذكور نظراً لضخامة العدد $n = 100$.

مثال (٤,١٠)

في أحد اختبارات الرياضيات طرح على سعيد 250 سؤالاً مكتوباً ، وأمام كل سؤال أربع إجابات واحدة منها تمثل الإجابة الصحيحة ، واختار سعيد 100 سؤالا منها ، وأجاب عليها بالتحزير (التخمين) نظراً لعدم تحضيره لهذه المادة بالشكل الجيد . ما هو احتمال أن يقود هذا التخمين إلى إجابات صحيحة عددها من 35 إلى 40 ؟

الحل

نفرض أن X يمثل عدد الإجابات الصحيحة بين 100 إجابة ، فيكون المطلوب حساب :

$$P[35 \le X \le 40] = \sum_{x=35}^{40} b\left(x; 100, \frac{1}{4}\right)$$

وباستخدام التقريب الطبيعى للتوزيع الحدانى بالوسيطين :

$$\mu = n.p = (100)\frac{1}{4} = 25$$

$$\sigma = \sqrt{npq} = \sqrt{(100)\left(\frac{1}{4}\right)\left(\frac{3}{4}\right)} = 4.629$$

نجد أن المساحة الواقعة بين النقطتين $x_1 = 34.5$ ، $x_2 = 40.5$ توافق المساحة بين القيمتين التاليتين للمتغير Z :

$$Z_1 = \frac{34.5 - 25}{4.629} = 2.05$$

$$Z_2 \frac{40.5 - 25}{4.629} = 3.348$$

وهكذا نجد أن :

$$P[35 \le X \le 40] \simeq P[2.05 < Z < 3.348]$$
$$= P[Z < 3.35] - P[Z < 2.05]$$

ومن الجدول IV نجد أن :

$$P[25 \le X \le 30] = 0.9996 - 0.9798$$
$$= 0.0198$$

(٤,٤) التوزيعات غما ، الأُسي ، وكاى – مربع
Gamma, exponential and chi-Square distribution

على الرغم من أن التوزيع الطبيعى يمكن أن يستخدم لحل الكثير من المسائل الهندسية والعلمية ، فإن هناك عدداً من الحالات تتطلب نوعا آخر من دوال الكثافة . من هذ الدوال ، هناك ما يسمى بالتوزيع غماً ، التوزيع الأسسى ، وأيضا التوزيع كاى مربع . سنبدأ فيما يلى بدراسة التوزيع غما .

لقد اشتق التوزيع غما اسمه من دالة معروفة بشكل جيد لدى الرياضيين تدعى بالدالة غما ، وقبل دراسة هذا التوزيع ، لابد من ذكر بعض الخواص الهامة لهذه الدالة .

تعريف (٤,٢) الدالة غما Gamma function

تعرف الدالة غما بأنها :

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} \cdot e^{-x} \cdot dx \qquad : \qquad x > 0$$

والخواص التى تحققها هذه الدالة نوردها على التتالى :

١ — $\Gamma(\alpha) = (\alpha - 1)\,\Gamma(\alpha - 1)$

٢ — $\Gamma(\alpha) = (\alpha - 1)\,!$ عندما تكون α رقما صحيحاً موجباً

٣ — $\Gamma(1) = 1$

٤ — $\Gamma(\frac{1}{2}) = \sqrt{\pi}$

ويمكن بسهولة برهان هذه الخواص . فمثلا لو كاملنا بالتجزئة التكامل الوارد فى التعريف (٤,٢) بأن نضع : $u = x^{\alpha-1}$, $dv = e^{-x}\ dx$ فإننا نجد أن :

$$\Gamma(\alpha) = -e^{-x} \cdot x^{\alpha-1} \Big|_0^\infty + \int_0^\infty e^{-x} (\alpha-1) \cdot x^{\alpha-2}\, dx$$

$$= (\alpha-1) \int_0^\infty e^{-x} \cdot x^{\alpha-2}\, dx$$

$$\Gamma(\alpha) = (\alpha-1)\,\Gamma(\alpha-1)$$

وهو برهان الخاصة الأولى .. ولو أعدنا مع التكرار ما ذكرناه لوجدنا أن :

$$\Gamma(\alpha) = (\alpha-1)\,.\,(\alpha-2)\,\Gamma(\alpha-2)$$
$$= (\alpha-1)(\alpha-2)\,.\,(\alpha-3)\,\Gamma(\alpha-3)$$
$$= (\alpha-1)(\alpha-2) \ldots\ldots \Gamma(1)$$

إلا أننا نلاحظ أن :

$$\Gamma(1) = \int_0^\infty e^{-x}\, dx = 1$$

وبذلك نحصل على :

$$\Gamma(\alpha) = (\alpha-1)\,!$$

وهكذا تتحقق الخاصتان ٢ ، ٣ أيضا . وبتعويض $x = \frac{1}{2} u^2$ فى التكامل :

فإننا نجد أن :

$$\Gamma\left(\frac{1}{2}\right) = \int_0^\infty e^{-x} \cdot x^{-\frac{1}{2}}\, dx$$

وكما نعلم فإن :

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{2} \int_0^\infty e^{-\frac{1}{2}u^2}\, du$$

وبذلك نعلم فإن :

$$\int_0^\infty e^{-\frac{1}{2}u^2}\, du = \frac{1}{2}\sqrt{2\pi}$$

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{2} \cdot \frac{1}{2}\sqrt{2\pi} = \sqrt{\pi}$$

وهو كما نرى برهان الخاصة الرابعة :

يتضمن تعريفنا للتوزيع غما الدالة غما .

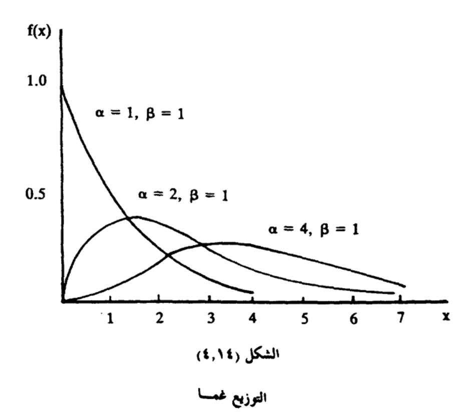
التوزيع غما Gamma distribution

نقول بأن للمتغير المستمر X توزيع غما بالوسيطين β , α إذا كانت كثافته من الشكل :

$$f(x) = \begin{cases} \frac{1}{\beta^{\alpha}\,\Gamma(a)} x^{\alpha-1} e^{-\frac{x}{\beta}} & : \quad x > 0 \\ 0 & : \quad \text{فيما عدا ذلك} \end{cases}$$

وبفرض أن كلا من $\alpha, \beta > 0$.

يوضح الشكل (٤,١٤) عدداً من توزيعات غما من أجل قيم متعددة لـ β و α . وفى الحالة الخاصة إذا كان $\alpha = 1$ ، فإننا نسمى التوزيع الناتج عن التوزيع غما بالتوزيع الأسي .



الشكل (٤,١٤)

التوزيع غمـا

التوزيع الأسي Exponential distribution

نقول بأن للمتغير المستمر X توزيعاً أُسياً بالوسيط β إذا كانت دالة كثافته من الشكل

$$f(x) = \begin{cases} \frac{1}{\beta} \cdot e^{-\frac{x}{\beta}} & : x > 0 \\ 0 & : \text{فيما عدا ذلك} \end{cases}$$

حيث إن $\beta > 0$. وللتوزيع الأسسى تطبيقات واسعة فى مجال الإحصاء وعلى الأخص فى نظرية الثقة ، وفى مواضع الارتال . نسوق فيما يلى أحد هذه التطبيقات الممتعة لهذا التوزيع .

مثال (٤,١١)

تحوى مجموعة نوعا خاصا من العناصر ، لكل منها عمر T (بقدر السنوات) ، ويتوزع وفقا للتوزيع الأسى بالمتغير $\beta = 6.5$. أخذت عشرة عناصر منها وركبت (نصبت) بأشكال مختلفة ، ما هو احتمال أن يستمر سبعة منها على الأقل في تأدية وظائفها بعد مرور تسعة أعوام على تركيبها ؟

الحل

نلاحظ أن احتمال أن يستمر عنصر ما من هذه العناصر فى تأدية وظيفته بعد مرور تسعة أعوام على تركيبه هو :

$$P(T > 9) = \frac{1}{6.5} \int_{9}^{\infty} e^{-\frac{t}{6.5}} dt = e^{-\frac{9}{6.5}} = 0.25$$

لنفرض أن X يمثل عدد العناصر التى استمرت فى تأدية وظائفها بعد مرور تسعة أعوام على تركيبها ، عندئذ باستخدام التوزيع الحدانى نجد أن :

$$P(X \geq 7) = \sum_{x=7}^{10} b(x; 10, 0.25)$$

$$= 1 - \sum_{x=0}^{6} b(x; 10, 0.25)$$
$$= 1 - 0.9965$$
$$= 0.0035$$

يدعى التوزيع الخاص الثانى الذى نحصل عليه من توزيع غما بوضع $\alpha = \frac{\nu}{2}$, $\beta = 2$ (حيث يمثل ν عدداً صحيحاً موجباً) بالتوزيع كاى مربع بـ ν درجة من الحرية .

التوزيع كاى – مربع Chi-Square distribution

يتوزع المتغير العشوائى المستمر X وفقا للتوزيع غما بـ درجة من الحرية إذا كانت دالة كثافته من الشكل :

$$f(x) = \begin{cases} \dfrac{1}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})} x^{\frac{\nu}{2}-1} \cdot e^{-\frac{x}{2}} & : x > 0 \\ 0 & : \text{فيا عدا ذلك} \end{cases}$$

حيث يمثل ν عدداً صحيحاً موجباً .

ويعتبر توزيع غما واحدا من الأدوات الرئيسية فى حقل اختبار الفرضيات الذى سندرسه فى الفصول القادمة . سنحاول فيما يلى إيجاد توقع وتباين كل من هذه المتغيرات العشوائية .

نظرية (٤,٣)

إن توقع وتباين متغير وايبل يعطيان بالعلاقتين :

$$\mu = \alpha.\beta \ , \ \sigma^2 = \alpha.\beta^2$$

البرهان

نلاحظ أن العزم من المرتبة r حول المبدأ تكتب بالشكل التالى :

$$\mu_r = E(X^r) = \frac{1}{\beta^\alpha . \Gamma(\alpha)} \int_0^\infty x^{r+\alpha-1} . e^{-\frac{x}{\beta}} dx$$

وبإجراء تغير فى المتحول بواسطة العلاقة $\frac{x}{\beta} = y.$ نجد أن :

$$\mu_r = \frac{\beta^r}{\Gamma(\alpha)} \int_0^\infty y^{r+\alpha-1} . e^{-y} dy$$

$$= \frac{\beta^r}{\Gamma(\alpha)} \Gamma(\alpha+r)$$

وهكذا فإن :

$$\mu = \mu_1^{'} = \frac{\beta \, \Gamma(\alpha+1)}{\Gamma(\alpha)} = \alpha . \beta$$

كما نجد أيضا :

$$\sigma^2 = \mu_2 - \mu^2 = \frac{\beta^2 \, \Gamma(\alpha+2)}{\Gamma(\alpha)} - \alpha^2 . \beta^2$$

$$= \alpha . \beta^2$$

نتيجة (٤,١)

إن توقع وتباين لمتغير الأسى هما :

$$\mu = \beta \ , \ \sigma^2 = \beta^2$$

نتيجة (٤,٢)

أن توقع وتباين المتغير كاى – مربع هما : $\mu = \nu \ , \ \sigma^2 = 2.\nu$

(٤,٥) توزيع وايبل Weibull distribution

تمكنت التكنولوجيا الحديثة من تصميم أنظمة عديدة معقدة تعتمد عملياتها على الوثوقية من العناصر المختلفة التى تتكون منها هذه الأنظمة . فمثلا يمكن للفيوز المستخدم فى جهاز معين أن يحترق ، كما يمكن للعمود المعدنى أن يلتوى ، ويمكن لجهاز حساس بالحرارة أن يتعطل . إن هذه العناصر المتماثلة والمعرضة لظروف بيئية معينة ستتلاشى فى أزمنة لا يمكن التنبؤ بها . إن عمر أى عنصر يقاس من فترة تكوينه إلى فترة فنائه ، وهذا الزمن هو متغير عشوائى مستمر T له كثافة احتمالية f(t) وإن أحد التوزيعات المستخدمة على نطاق واسع والذى يتعامل مع مثل هذه المشاكل كالوثوقية واختبار العمر يدعى بتوزيع وايبل .

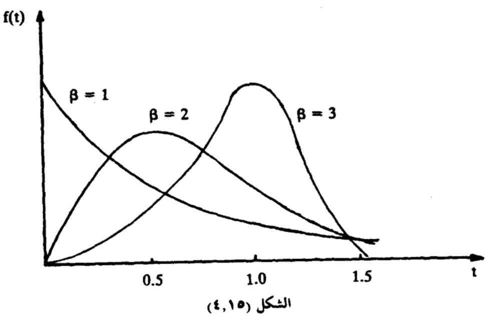
توزيع وايبل

المتغير العشوائى المستمر T له توزيع وايبل بالمتغيرين β ، α إذا كانت كثافته الاحتمالية من الشكل :

$$f(t) = \begin{cases} \alpha . \beta . t^{\beta - 1} e^{-\alpha . t^{\beta}} & : t > 0 \\ 0 & \text{فيما عدا ذلك:} \end{cases}$$

حيث $\alpha > 0 , \beta > 0$

ولكثافة وايبل منحنيات متعددة بحسب القيم المختلفة لكل من α , β . وفى الشكل (٤,٥) نرى أن منحنى الكثافة يتغير بشكل كبير بتغير قيم الوسطاء α , B وعلى الأخص بتغير قيم β . نلاحظ أن توزيع وايبل يصبح توزيعاً أسيا إذا كان $\beta = 1$ ومن أجل قيم $\beta > 1$ يصبح شكل منحنى الكثافة قريبا من المنحنى الجرسى ويشبه المنحنى الطبيعى ، ويبدو عدم تناظره للعيان .



الشكل (٤,١٥)

نظرية (٤,٣)

إن توقع وتباين متغير وايبل يعطيان بالعلاقتين :

$$\mu = \alpha^{-\frac{1}{\beta}} \Gamma\left(1 + \frac{1}{\beta}\right)$$

$$\sigma^2 = \alpha^{-\frac{2}{\beta}} \left\{ \Gamma\left(1 + \frac{2}{\beta}\right) - \left[\Gamma\left(1 + \frac{1}{\beta}\right) \right]^2 \right\}$$

ولتطبيق توزيع وايبل فى نظرية الوثوقية ، نعرف أولاً وثوقية عنصر ، كاحتمال أن يكون دالة لزمن محدد على الأقل تحت ظروف تجريبية محددة . فبفرض أن R(t) يمثل وثوقية عنصر ما فى زمن t ، فإننا نستطيع أن نكتب عندئذ :

$$R(t) = P(X > t)$$

$$= \int_t^{\infty} f(t)\, dt$$

$$= 1 - F(t)$$

حيث يمثل F(t) التوزيع التراكمى للمتغير T . وإن الاحتمال الشرطى لفناء عنصر فى مجال

زمنى من $T = t$ إلى $T = t + \Delta t$ يعطى بالعلاقة :

$$\frac{F(t + \Delta t) - F(t)}{R(t)}$$

وبتقسيم هذه النسبة على Δt ، ويأخذ النهاية عند ما تسعى Δt إلى الصفر ، فإننا نحصل على معدل الفشل ، والذى نرمز له بالرمز $Z(t)$. لذلك فإن :

$$Z(t) = \lim_{\Delta t \to 0} \frac{F(t + \Delta t) - F(t)}{\Delta t} \cdot \frac{1}{R(t)}$$

$$= \frac{F'(t)}{R(t)} = \frac{f(t)}{R(t)} = \frac{f(t)}{1 - F(t)}$$

والعلاقة :

$$Z(t) = \frac{f(t)}{1 - F(t)}$$

تعبر عن معدل الفشل فى صيغة توزيع زمن الفشل .

وبما أن $R(t) = 1 - F(t), \acute{R}(t) = -\acute{F}(t)$ ، فإننا نستطيع أن نكتب المعادلة التفاضلية التالية :

$$Z(t) = \frac{-\acute{R}(t)}{R(t)} = \frac{-d\,[\ln R(t)]}{dt}$$

وبحلها نجد أن :

$$\ln R(t) = -\int Z(t)\,dt$$

أو :

$$R(t) = e^{-\int Z(t)dt} + C$$

مثال (٤,١٢)

بين أن معدل الفشل يعطى بالعلاقة التالية :

$$Z(t) = a.\beta \,.\, t^{\beta - 1} \quad : t > 0$$

إذا وفقط إذا كان زمن الفشل يتوزع وفق توزيع وايبل

$$f(t) = a.\beta.t^{\beta-1}.e^{-at^{\beta}} : t > 0$$

الحل

الحالة الأولى :

نفرض أن :

$$Z(t) = a.\beta.t^{\beta-1} : t > 0$$

نلاحظ أن :

$$R(t) = e^{-\int z(t)dt} = e^{-\int \alpha.\beta t^{\beta-1} dt} = e^{-\alpha.t^{\beta}} + C$$

ومن كون أن $R(0) = 1$ ، نجد أن $C = 0$ لذلك فإن :

$$R(t) = e^{-t^{\beta}}$$

كما نجد أيضا أن :

$$f(t) = a.\beta.t^{\beta-1}.e^{-\alpha t^{\beta}} : t > 0$$

الحالة الثانية :

نفرض أن :

$$f(t) = \alpha.\beta.t^{\beta-1} e^{-\alpha.t^{\beta}} : t > 0$$

نعلم أن :

$$Z(t) = \frac{f(t)}{R(t)}$$

حيث يمثل :

$$R(t) = 1 - F(t) = 1 - \int_0^t \alpha.\beta.x^{\beta-1}.e^{-\alpha.x^{\beta}} dx$$

$$= 1 + \int_0^t d(e^{-\alpha.x^{\beta}})$$

$$= e^{-\alpha.t^{\beta}}$$

وهكذا نجد أن :

$$Z(t) = \frac{\alpha.\beta.t^{\beta-1}.e^{-\alpha t^{\beta}}}{e^{-\alpha.t^{\beta}}} \alpha.\beta.t^{\beta-1} : t > 0$$

تمارين محلولة

تمرين (١)

بفرض X متغير عشوائى طبيعى بالمتغيرين $\mu = 30, \sigma = 4$. أوجد :

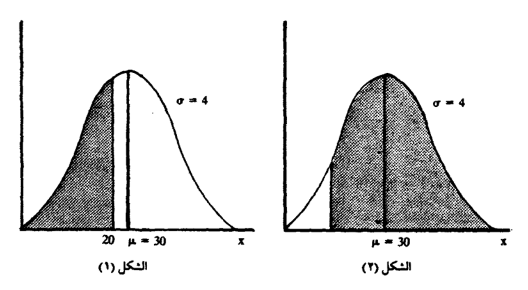
(١) المساحة الواقعة أسفل النقطة 20 .

(٢) المساحة الواقعة أعلى النقطة 25 .

(٣) المساحة الواقعة بين النقطتين 33 ,42 .

(٤) النقطة التى تحصر قبلها مساحة قدرها 45% .

(٥) النقطة التى تحدد بعدها مساحة قدرها 13% .



الشكل (١) الشكل (٢)

الحل

من الشكل (١) نلاحظ أن المساحة الواقعة أسفل النقطة 20 تمثل مساحة القسم المظلل من الشكل لذلك فإن :

$$P[X < 20] = P\left[\frac{X-\mu}{\sigma} < \frac{20-30}{4}\right]$$

$$= P\left[Z < -\frac{10}{4}\right]$$

$$= P[Z < -2.5]$$

حيث يمثل Z متغيراً طبيعياً معيارياً . وبالعودة إلى الجدول IV نجد أن :

$$P[X < 20] = 0.0062$$

كذلك فإن المساحة الواقعة أعلى النقطة 25 تمثل مساحة القسم المظلل على الشكل (٢) وهذه المساحة تحسب بالعلاقة التالية :

$$P[X < 25] = 1 - P[X < 25]$$

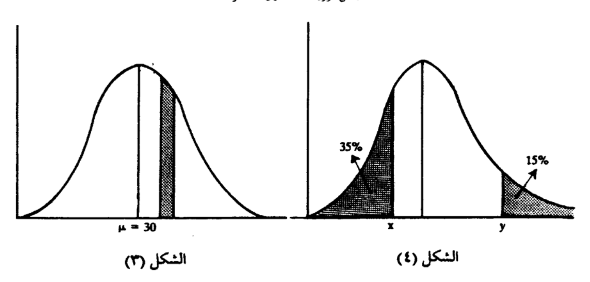
$$= 1 - P\left[Z < -\frac{25-30}{4}\right]$$

$$= 1 - P\left[Z < -\frac{-5}{4}\right]$$

$$= 1 - P[Z < -1.25]$$

$$= 1 - 0.1056$$

$$= 0.8944$$



الشكل (٣) الشكل (٤)

يوضح القسم المظلل على الشكل (٣) المساحة المحصورة بين النقطتين 33 ,42 ، وهذه المساحة تحسب على النحو التالى :

$$P\,[33 < X < 42] = P\left[\frac{33-30}{4} < Z < \frac{42-30}{4}\right]$$

$$= P\,[0.75 < Z < 3]$$

$$= P[Z < 3] - P[Z < 0.75]$$

$$= 0.9987 - 0.7734$$

$$= 0.2253$$

كما أن النقطة x التى تحصر قبلها مساحة قدرها 35% موضحة على الشكل (٤) . يمثل القسم المظلل على الشكل المساحة المذكورة .

رابعاً : نجد أن النقطة التى تحصر قبلها مساحة قدرها 35% هى النقطة x المحققة للعلاقة :

$$P[X < x] = 0.35$$

ومنه :

$$P\left[Z < \frac{x-30}{4}\right] = 0.35$$

ونلاحظ من الجدول IV أنه يقابل المساحة 0.35 النقطة 0.385 - = $Z = \frac{x-30}{4}$

ولذلك فإن :

$$x = (-0.385)(4) + 30$$

ومنه :

$$x = 28.46$$

وأخيراً فإن النقطة التى تحدد بعدها مساحة قدرها 15% هى النقطة x المحققة للعلاقة :

$$P[X > x] = 0.15$$

$$P\left[Z > \frac{x-30}{4}\right] = 0.15$$

ومنه :

$$1 - P\left[Z < \frac{x-30}{4}\right] = 0.15$$

ولذلك فإن :

$$P\left[Z < \frac{x-30}{4}\right] = 1 - 0.15$$

ومن الجدول IV نجد أن النقطة التي تحدد لنا مساحة قدرها 0.85 هى النقطة 1.035 = Z

ومنه :

$$Z = \frac{x-30}{4} = 1.035$$

ولذلك فإن :

$$x = 4 . (1.035) + 30$$

$$x = 34.14$$

تمرين (٢)

Y متغير عشوائى طبيعى بالوسيطين $\mu = 18, \sigma = 2.5$. حدد القيمة k بحيث تتحقق المساواة :

$$P[X > k] = 0.1539$$

الحل

نلاحظ أن :

$$P\left[\frac{X - \mu}{\sigma} > \frac{k - 18}{2.5}\right] = 0.1539$$

$$P\left[Z > \frac{k - 18}{2.5}\right] = 0.1539$$

ومنه :

$$1 - P\left[Z < \frac{k - 18}{2.5}\right] = 0.1539$$

ولذلك فإن :

$$P\left[Z < \frac{k - 18}{2.5}\right] = 1 - 0.1539 = 0.8461$$

هذا ، ويقابل المساحة 0.8461 فى الجدول IV النقطة :

$$Z = \frac{k - 18}{2.5} = 1.02$$

ولذلك فإن :

$$k = 20.55$$

تمرين (٣)

أوجد الخطأ المرتكب فى حساب $\sum_{x=1}^{4} b(x; 20, 0.1)$ بواسطة التقريب بالتوزيع الطبيعى .

الحل

من الواضح أن :

$$\sum_{x=1}^{4} b(x; 20; 0.1) = \sum_{x=0}^{4} b(x; 20, 0.1) - b(0; 20, 0.1)$$

من الجدول II نجد أن :

$$= 0.9568 - 0.1216$$

$$= 0.8352$$

وحسب التقريب بالتوزيع الطبيعى نجد أن :

$$\sum_{x=1}^{4} b(x; 20, 0.1) = P[1 \le X \le 4]$$

ونلاحظ أن :

$$n = 20 , p = 0.1$$

ولذلك فإن :

$$\sigma = \sqrt{np.q} = 1.34 \quad , \quad \mu = np = 2$$

كما أن :

$$P[1 \le X \le 4] = P[0.5 < X < 4.5]$$

$$= P\left[\frac{0.5-2}{1.34} < Z < \frac{4.5-2}{34}\right]$$

$$= P\left[Z < \frac{4.5}{1.34}\right] - P\left[Z < \frac{-1.5}{1.34}\right]$$

$$= P[Z < 1.869] - P[Z < -1.119]$$

$$= 0.9686 - 0.1314 = 0.8372$$

والخطأ المرتكب هو :

$$0.8373 - 0.8352 = 0.0021$$

تمرين (٤)

إذا علمت أن احتمال شفاء إنسان من مرض القلب نتيجة إجراء عملية جراحية له هو 0.9 . ما هو احتمال شفاء من 84 إلى 95 مريضاً من أصل مئة أجريت لهم نفس العملية ؟

الحل

نفرض أن X يمثل عدد المرضى الذين تم شفاؤهم من أصل مئة مريض . نلاحظ أن للمتغير X توزيعاً حدانياً فى تجربة فيها $n = 100$ ، $p = 0.9$ ، $\mu = n \cdot p = 90$ ، $\sigma = 3$. نجد أن :

$$P[84 \le X \le 95] = \sum_{x=84}^{95} b(x; 100, 0.9)$$

$$= P[83.5 < X < 95.5]$$

$$= P\left[\frac{83.5-90}{3} < Z < \frac{95.5-90}{3}\right]$$

$$= P[Z < 1.83] - P[Z < -2.166]$$

$$= 0.9664 - 0.0150$$

$$= 0.9514$$

تمرين (٥)

متغير عشوائى X له توزيع غما بالمتغيرين $\beta = 1, \alpha = 2$ أوجد :

$$P[1.8 < X < 2.4]$$

الحل

من المعلوم أن الكثافة الاحتمالية للمتغير X فى هذه الحالة هي :

$$f(x) = \frac{1}{1^2 . \Gamma(2)} x^{2-1} . e^{-\frac{x}{1}} : x > 0$$

وبما أن $1 = (2)$ لذلك :

$$f(x) = \begin{cases} x . e^{-x} & : x > 0 \\ 0 & \text{: فيما عدا ذلك} \end{cases}$$

وبما أن :

$$P[1.8 < X < 2.4] = \int_{1.8}^{2.4} f(x)\, dx$$

لذلك فإن :

$$= \int_{1.8}^{2.4} x . e^{-x}\, dx$$

وبالتكامل بالتجزئة ، وبفرض أن $U = x$, $dv = e^{-x}\,dx$ نجد أن :

$$= -e^{-x}(x+1)\Big|_{x=1.8}^{2.4}$$

$$= -0.30844 + 0.462836$$

$$= 0.154396$$

تمرين (٦)

إذا علمت أن الاستهلاك اليومى للماء فى مدينة معينة هو متغير عشوائى له تقريباً توزيع غما بالوسيطين $\alpha = 2, \beta = 3$ (علماً بأن هذا الاستهلاك بملايين اللترات) . فإذا كان مخزون الماء اليومى فى هذه المدينة 9 مليون لتر ، فما هو احتمال نفاذ الماء فى أحد الأيام ؟

الحل

لنرمز للاستهلاك اليومى للماء فى هذه المدينة بالرمز X ، فيكون لهذا المتغير توزيعاً قريبا من التوزيع غما بالوسيطين $\alpha = 2,\beta = 3$ وهذا يعنى أن :

$$f(x) = \begin{cases} \dfrac{1}{3^2 \cdot \Gamma(2)}\, x^{2-1}\, e^{-\frac{x}{3}} & : x > 0 \\ 0 & : \text{فيما عدا ذلك} \end{cases}$$

والاحتمال المطلوب هو :

$$P[X \geq 9] = 1 - P[X < 9]$$

$$= 1 - \int_0^9 f(x)\,dx$$

$$= 1 - \int_0^9 \frac{1}{9}\, x \cdot e^{-\frac{x}{3}}\,dx$$

وبإجراء تغيير فى المتحول بالعلاقة $U = \frac{x}{3}$ نجد أن :

$$P[x \geq 9] = 1 - \int_0^3 U e^{-U} dU$$

وبالمكاملة بالتجزئة نجد أن :

$$= 1 + \int_0^3 U d(e^{-U})$$

$$= 1 + U e^{-U} \Big|_0^3 - \int_0^3 e^{-U} dU$$

$$= 1 + 3e^{-3} + e^{-U} \Big|_0^3$$

$$= 1 + 3e^{-3} + e^{-3} - 1$$

$$= 4e^{-3}$$

$$= 0.1992$$

تمرين (٧)

إذا كان عمر نوع معين من الأزرار الكهربائية (بالأعوام) يتوزع وفقا للتوزيع الأسى . بمعدل فشل قدره $\beta = 2$. ركبنا مئة زر كهربائى من هذه الأزرار بأشكال مختلفة . ما هو احتمال أن يتعطل على الأكثر ثلاثين زراً من هذا الأزرار خلال العام الأول ؟

الحل

نرمز لعدد الأزرار المعطلة من بين المئة المركبة بـ X .

كما نرمز بـ Y لعمر كل زر بالسنوات

نلاحظ أن لـ Y كثافة احتمالية قدرها :

$$f(y) = \begin{cases} \frac{1}{2} e^{-\frac{y}{2}} & : y > 0 \\ 0 & \text{: فيما عدا ذلك} \end{cases}$$

أما احتمال أن يتعطل أى زر خلال عام من تركيبه فهو يساوى :

ومنه :

$$P[Y \leq 1] = \int_0^1 \frac{1}{2} e^{-\frac{y}{2}} dy$$

$$p = 0.3935$$

وذلك فإن :

$$q = 0.6065$$

وهكذا نجد أن المتوسط والانحراف المعيارى للمتغير X هما :

$$\mu = n.p = 39.34 \ , \ \sigma = 4.885$$

وهكذا نجد أن الاحتمال المطلوب هو :

$$P[X \leq 30] = P[X < 30.5]$$

$$= P\left[Z < \frac{30.5 - 39.34}{4.885} \right]$$

$$= P[Z < -1.81]$$

من الجدول IV نجد أن :

$$= 0.0352$$

تمارين عامة

(١) متغير عشوائى طبيعى وسطاؤه $\mu = 200$ و $\sigma^2 = 100$ أوجد :

أ – المساحة الواقعة تحت المنحنى الطبيعى وأعلى النقطة 179

ب – المساحة الواقعة تحت المنحنى الطبيعى وأسفل النقطة 214

(٢) باستخدام جدول التوزيع الطبيعى احسب الاحتمالات التالية :

$$P[-1 < Z < 1]\ ,\ P[-0.9 < Z < 0]\ ,\ P[Z > -0.9]$$

(٣) يتوزع عمر نوع من التلفزيونات وفقا للتوزيع الطبيعى بالوسط $\mu = 3.1$ سنة والانحراف المعيارى 1.2 سنة . فإذا كانت الشركة تكفل كل تلفزيون لمدة سنة . فما هى نسبة التلفزيونات المباعة التى ستضطر الشركة لاستبدالها بتلفزيون جديد ؟

(٤) إذا علمت أن عدد الصفقات التجارية التى يقوم بها التاجر محمد يتبع التوزيع الحدانى ، وإذا علمت أن احتمال اتمام صفقة مع زبون عند قدومه هو 0.4 . فما هو احتمال ألا يقل عدد صفقاته التى يتمها عن العدد 25 صفقة إذا استقبل مئة زبون فى يوم معين ؟

(٥) استخدم جدول χ^2 لحساب الاحتمالات التالية :

أ – $P[\chi^2 \geq 30.578]$ من أجل $\nu = 15$

ب – $P[\chi^2 \geq 130844]$ من أجل $\nu = 26$

جـ – $P[\chi^2 < 0.484]$ من أجل $\nu = 4$

د – $P[\chi^2\ 20.278]$ من أجل $\nu = 7$

هـ – $P[4.575 < \chi^2 < 24.725]$ من أجل $\nu = 11$

(٦) إذا علمت أن جملة من الملاحظات تتوزع وفقا للتوزيع الطبيعى . ما هى النسبة المئوية لاختلاف هذه الملاحظات عن المتوسط بـ

أ – أكثر من 1.3σ

ب – أقل من 0.52σ

(٧) إذا علمت أن معدل سقوط الأمطار فى مدينة ما فى شهر مارس (آذار) هو 9.22 سم مقربا لأقرب مئة من السنتيمتر . وبفرض أن سقوط الأمطار يتوزع وفقا للتوزيع الطبيعى بالانحراف المعيارى 2.82 سم . ما هو احتمال أن يهطل فى مارس القادم فى نفس المدينة .

أ – أقل من 1.84 سم من المطر .
ب – أكثر من 5 سم وليس أكثر من 7 سم .
جـ – أكثر من 13.8 سم .

(٨) احسب الخطأ المرتكب لدى تقريب $\sum_{x=2}^{14} b(x; 8, 0.1)$ بواسطة المنحنى الطبيعي .

(٩) أوجد التوقع والتباين لتوزيع وايبل .

(١٠) نقول بأن للمتغير العشوائى X توزيعا من نوع بتا بالوسيطين α و β إذا كانت دالة كثافته من الشكل :

$$f(x) = \begin{cases} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,.\,\Gamma(\beta)} x^{\alpha-1} \,.\, (1-x)^{\beta-1} & : 0 < x < 1 \\ 0 & : \text{فيما عدا ذلك} \end{cases}$$

حيث إن :

$$\alpha > 0 \,,\, \beta > 0$$

فإذا كانت النسبة المئوية لماركة معينة من التلفزيونات التى تتطلب التصليح خلال العام الأول لعملها تمثل متغيراً عشوائياً له توزيع بتا بالوسيطين ، فأوجد احتمال أن يكون 80% من التلفزيونات المباعة هذا العام من نفس الماركة تتطلب الإصلاح خلال السنة الأولى لإنتاجها .



الفصل الخامس

دوال المتغيــرات العشوائية

■ تغير المتغيرات ■ الدوال المولدة للعزوم ■ العينة العشـــوائية

■ نظرية العينات ■ توزيع المعاينة للوسط ■ توزيع المعاينة للمتغير العشوائي

$\frac{(n-1)S^2}{\sigma^2}$ ■ التـــوزيع t ■ التـــوزيع F ■ تمـــارين محلـــولة

■ تمارين عامة .



(٥,١) تغيير المتغيرات Change of variables

نحتاج فى أغلب الأحيان إلى معرفة التوزيع الاحتمالى لدالة بمتغير واحد أو أكثر . فبفرض أن X يمثل (متغيراً عشوائياً منقطعاً بالتوزيع الاحتمالى $f(x)$ ، وأن $Y = u(X)$ يمثل تطبيقاً واحداً لواحد بين مجموعة قيم X ومجموعة قيم Y) . نحن نرغب فى إيجاد التوزيع الاحتمالى للمتغير الجديد Y . ومن المهم أن نلاحظ أن التطبيق واحد لواحد الجديد يضمن لنا مقابلة كل قيمة x لـ X بقيمة وحيدة y لـ Y والعكس بالعكس .

ويبدو لنا واضحا من خلال مناقشتنا للتوزيعات الاحتمالية المنقطعة فى الفصل الثانى أن المتغير العشوائى Y يفترض القيمة y عندما يفترض المتغير X القيمة $\lambda(y)$ (حيث يمثل $\lambda(y)$ الحل العكسى للمعادلة $(Y = u(x))$. وهكذا نجد أن التوزيع الاحتمالى للمتغير Y يعطى بالعلاقة :

$$h(y) = P(Y = y) = P[X = \lambda(y)] = f[\lambda(y)]$$

نظرية (٥,١)

بفرض أن X يمثل متغيراً عشوائياً منقطعاً بالتوزيع الاحتمالى $f(x)$ ، وأن $Y = u(x)$ يمثل تطبيقاً واحداً لواحد بين مجموعة قيم X ومجموعة قيم Y بحيث تكون للمعادلة $Y = u(x)$ حلا عكسيا من الشكل $X = \lambda(y)$. عندئذ يكون التوزيع الاحتمالى للمتغير Y :

$$h(y) = f[\lambda(y)]$$

مثال (٥,١)

نفرض أن X متغير عشوائى هندسى بالتوزيع الاحتمالى :

$$f(x) = \frac{2}{5} \cdot \left(\frac{3}{5}\right)^{x-1} \quad ; x = 1, 2, 3, \ldots$$

ولنفتش الآعن توزيع الاحتمال للمتغير $Y = x^3$

الحل

بما أن جميع قيم المتغير X موجبة ، فالتطبيق Y هو تطبيق واحد لواحد بين مجموعة قيم X ومجموعة قيم Y . لذلك حسب النظرية (٥,١) والعلاقة $y = x^3$ التي تعطي $X = Y^{1/3}$ نجد أن توزيع الاحتمال للمتغير Y هو :

$$h(y) = f[\sqrt[3]{y}] = \begin{cases} \left(\frac{2}{5}\right)\left(\frac{3}{5}\right)^{y^{\frac{1}{3}}-1} & : y = 1, 8, 27, \ldots \\ = 0 & : \text{فيما عدا ذلك} \end{cases}$$

لنفرض الآن لدينا متغيرين عشـــوائيين منقطعين X_1 , X_2 وتوزيعهما الاحتمالى المشـــترك هو $f(x_1, x_2)$. نحن نرغب فى إيجاد التوزيع الاحتمالى المشترك $g(y_1 , y_2)$ لمتغيرين عشوائيين جديدين $Y_1 = U_1(X_1, X_2), Y_2 = U_2(X_1, X_2)$ يحدد كلا منهما تقابلا واحدا لواحد بين مجموعة النقاط (x_1, x_2) ومجموعة النقاط (y_1 , y_2) . بحل المعادلتين :

$Y_2 = U_2(x_1, x_2), y_1 = U_1(x_1, x_2)$ فى وقت واحد بالنسبة لـ :

x_1, x_2 نجد أن : $x_2 = \lambda_2(y_1, y_2)\ x_1 = \lambda_1(y_1, y_2)$.

لذلك فإن المتغيرين Y_1, Y_2 سيفرضان على الترتيب القيمتين y_1, y_2 عندما يفترض المتغيران X_1, X_2 القيمتين $\lambda_2(y_1, y_2)$ و $\lambda_1(y_1, y_2)$ على الترتيب أيضا . ولذلك فإ التوزيع الاحتمالى المشترك لـ Y_1, Y_2 هو :

$$h(y_1, y_2) = P[Y_1 = y_1, Y_2 = y_2]$$

$$= P[X_1 = \lambda_1(y_1, y_2), X_2 = \lambda_2(y_1, y_2)]$$

$$= f[\lambda_1(y_1, y_2), \lambda_2(y_1, y_2)]$$

نظرية (٥,٢)

بفرض أن X_1, X_2 متغيران عشـوائيان منقطعان توزيعهما الاحتمالى المشـترك $f(y_1, y_2)$ وأن كلا من : $Y_1 = U_1(X_1, X_2)$ و $Y_2 = U_2(X_1, X_2)$ يمثلان تقابلا واحداً لواحد بين مجموعة النقاط (x_1, x_2) و (y_1, y_2) بحيث يكون للمعادلتين $Y_1 = U_1(y_1, y_2)$ و $Y_2 = U_2(y_1, y_2)$ حلين وحيدين بالنسبة لـ x_1, x_2 مثلا $x_2 = \lambda_2(y_1, y_2), x_1 = \lambda_1(y_1, y_2)$ عندئذ يكون التوزيع الاحتمالى المشترك لـ Y_1, Y_2 من الشكل :

$$h(y_1, y_2) = f[\lambda_1(y_1, y_2) \, , \, \lambda_2(y_1, y_2)]$$

والنظـرية (٥,٢) مفيدة إلى أبعد حد فى إيجاد التوزيع الاحتمالى لبـض المتغيرات مثـلا ، $Y_1 = U_1(X_1, X_2)$ ، وذلك فى الحالة التى يكون فيها المتغيران X_1, X_2 منقطعان ، وبالتوزيع الاحتمالى المشترك $f(x_1, x_2)$. ففى هذا نبحث ، قبل كل شيء ، عن توزيع الاحتمال المشترك $h(y_1, y_2)$. فيكون توزيع Y_1 هو التوزيع الهامشى من الدالة $h(y_1, y_2)$ ونجده بالجمع فوق جميع قيم y_2 ، فإذا رمزنا لتوزيع Y_1 بالرمز $h(y_1)$ لوجدنا أن :

$$h(y_1) = \sum_{y_2} h(y_1, y_2)$$

مثال (٥,٢)

بفرض أن X_1, X_2 يمثلان متغيرين عشوائيين مستقلان لهما توزيعين بواسونيين بالوسيطين μ_1, μ_2 على الترتيب . لنبحث عن توزيع الاحتمال المشـترك للمتغير الجديد $Y_1 = X_1 + X_2$

الحل

من الاستقلال العشوائى للمتغيرين X_1, X_2 نجد أن :

$$f(x_1, x_2) = f(x_1) \, . \, f(x_2)$$

ومنه :

$$f(x_1 \, , \, x_2) = \frac{e^{-\mu_1} \, . \, \mu_1^{x_1} \, . \, e^{-\mu_2} \, . \, \mu_2^{x_2}}{x_1 ! \quad x_2 !}$$

$$= \frac{e^{-(\mu_1 + \mu_2)} \cdot \mu_1^{x_1} \cdot \mu_2^{x_2}}{X_1 ! \ X_2 !}$$

حيث أن :

$$x_1 = 0, 1, 2, \ldots \quad , \quad x_2 = 0, 1, 2, \ldots$$

لنعرف الآن متغيراً عشوائياً جديدا مثلا $Y_2 = X_2$ ، فتكون الحلول للمعادلتين $Y_2 = X_2$ و $Y_1 = X_1 + X_2$ من الشكل $X_1 = Y_1 - Y_2, X_2 = y_2$ وباستخدام النظرية (٥,٢) نجد أن التوزيع الاحتمالى المشترك لكل من Y_1, Y_2 هو :

$$h(y_1 , y_2) = \frac{e^{-(\mu_1 + \mu_2)} \cdot \mu_1^{y_1 - y_2} \cdot \mu_2^{y_2}}{(y_1 - y_2) ! \ y_2 !}$$

حيث :

$$y_1 = 0, 1, 2, \ldots , \qquad Y_2 = 0, 1, 2, \ldots$$

وهكذا نجد أن التوزيع الهامشى لـ Y_1 هو :

$$h(y_1) = \sum_{y_2=0}^{y_1} h(y_1 , y_2)$$

$$= e^{-(\mu_1 + \mu_2)} \cdot \sum_{y_2=0}^{y_1} \frac{\mu_1^{y_1 - y_2} \cdot \mu_2^{y_2}}{(y_1 - y_2)! \ y_2 !}$$

$$= \frac{e^{-(\mu_1 + \mu_2)}}{y_1 !} \sum_{y_2=0}^{y_1} \frac{y_1 !}{y_2 ! \ (y_1 - y_2) !} \mu_1^{y_1 - y_2} \cdot \mu_2^{y_2}$$

$$= \frac{e^{-(\mu_1 + \mu_2)}}{y_1 !} \sum_{y_2=0}^{y_1} \binom{y_1}{y_2} \cdot \mu_1^{y_1 - y_2} \cdot \mu_2^{y_2}$$

وبملاحظة أن المجموع الأخير يمثل المقدار $(\mu_1 + \mu_2)^{y_1}$ فإننا نجد أن :

$$h(y_1) = \frac{e^{-(\mu_1 + \mu_2)} \ (\mu_1 + \mu_2)^{y_1}}{y_1 !} : y_1 = 0, 1, 2, \ldots.$$

من هذا نستنتج أن مجموع متغيرين عشوائيين بواسونيين مستقلين بالوسيطين μ_1 , μ_2 هو من جديد متغير بواسونى وسيطه يساوى مجموع الوسيطين أى $\mu_1 + \mu_2$.

ولإيجاد التوزيع الاحتمالى للمتغير العشوائى $Y = U(X)$ عندما يكون المتغير X مستمراً ، والتقابل U واحد لواحد فإننا نسوق النظرية التاليه :

نظرية (٥,٣)

ليكن X متغيراً عشوائياً مستمراً بالتوزيع الاحتمالى $f(x)$ ، وليكن $Y = U(X)$ تقابلاً واحداً لواحد بين مجموعة قيم X ومجموعة قيم Y بحيث يكون للمعادلة $y = u(x)$ حلاً عكسياً وحيداً $x = \lambda(y)$. عندئذ يكون توزيع الاحتمال للمتغير Y من الشكل :

$$g(y) = f[\lambda(y)] \, . \, |J|$$

حيث يمثل $J = \acute{\lambda}(y)$ ويدعى جاكوبى التقابل .

البرهان :

الحالة الأولى

لنفرض أن $y = U(x)$ يمثل دالة متزايدة كما هو موضح على الشكل (٥,١) . نلاحظ أنه إذا وقع Y فى المجال (a, b) فعندئذ سيقع المتغير X فى المجال $(\lambda(a), \lambda(b))$ لذلك فإن :

$$P[a < Y < b] = P\,[\lambda(a) < X < \lambda(b)]$$

$$= \int_{\lambda(a)}^{\lambda(b)} f(x)\, dx$$

ويتغير المتحول فى التكامل من x إلى y بالعلاقة $x = \lambda(y)$ ، وبملاحظة أن $dx = \lambda(y)\, dy$ فإننا نجد أن :

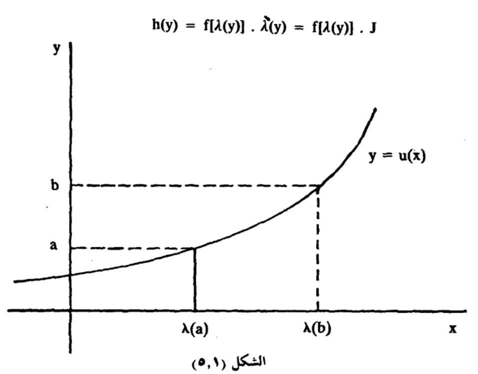
$$P\,[a < Y < b] = \int_a^b f\,[\lambda(y)] \, . \, \acute{\lambda}(y)\, dy$$

ومن المعلوم أن :

$$P[a < Y < b] = \int_a^b h(y)\, dy$$

وبمقارنة التكاملين السابقين نجد :

$$h(y) = f[\lambda(y)] \cdot \lambda'(y) = f[\lambda(y)] \cdot J$$



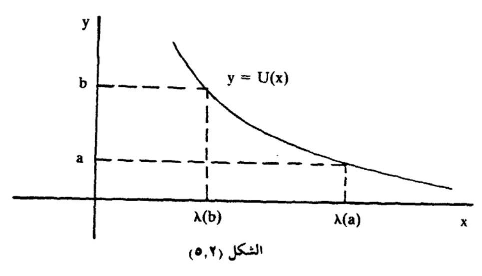
الشكل (٥,١)

وإذا لاحظنا أن $J = \lambda'(y)$ يمثل ميل المماس لمنحنى دالة متزايدة $y = U(X)$ ، لظهر لنا أن $J = |J|$ ولذلك فإن :

$$h(y) = f[\lambda(y)] \cdot |J|$$

الحالة الثانية

نفرض أن الدالة $y = U(x)$ متناقصة كما هو موضح على الشكل (٥,٢)



الشكل (٥,٢)

عندئذ نلاحظ أن :

$$P[a < Y < b] = P[\lambda(b) < X < \lambda(a)]$$

$$= \int_{\lambda(b)}^{\lambda(a)} f(x) \ dx$$

وإذا غيرنا المتحول من x إلى y فإننا نجد أن :

$$P[a < Y < b] = \int_{b}^{a} f[\lambda(y)] \ . \ \lambda'(y) \ dy$$

$$= - \int_{a}^{b} f[\lambda(y)] \ . \ \lambda(y) \ dy$$

من هذا نستنتج أن :

$$h(y) = - f[\lambda(y)] \ . \ \lambda'(y) = - f[\lambda(y)] \ . \ J$$

وفى هذه الحالة نعلم أن ميل المماس سالب ومنه $J = -|J|$ لذلك فإن :

$$h(y) = f[\lambda(y)] \ . |J|$$

مثال (٥,٣)

بفرض X متغير عشوائى مستمر بالتوزيع الاحتمالى :

$$f(x) = \begin{cases} \frac{3}{7} x^2 & : 1 < x < 2 \\ 0 & : \text{فيما عدا ذلك} \end{cases}$$

ولنفتش عن توزيع الاحتمال للمتغير $Y = X + 3$

الحل

إن الحل العكسى للمعادلة $y = x + 3$ هو $x = y - 3$. ومنه نستنتج أن $dx =$

dy . باستخدام النظرية (٥,٣) نجد أن الكثافة الاحتمالية للمتغير Y هى :

$$h(y) = \frac{3}{7} . (y - 3)^2$$

ومنه :

$$h(y) = \begin{cases} \frac{3}{7} . (y-3)^2 & : +4 < y < 5 \\ 0 & : \text{فيما عدا ذلك} \end{cases}$$

لإيجاد التوزيع الاحتمالى المشترك للمتغيرين $Y_1 = U_1(X_1, X_2)$، $Y_2 = U_2(X_1, X_2)$، وذلك فى الحالة التى يكون فيها المتغيران X_1, X_2 مستمرين ، ويكون كل من التقابلين U_2 ، U_1 ممثلا لتقابل واحد لواحد ، فإننا نحتاج إلى النظرية الإضافية التالية والتى سنوردها بدون برهان .

نظرية (٥,٤)

بفرض أن X_1, X_2 يمثلان متغيرين مستمرين بالتوزيع المشترك $f(x_1, x_2)$ ، وأن كلا من $Y_1 = U_1(X_1, X_2)$ ، $Y_2 = U_2(X_1, X_2)$ يمثلا تقابلا واحدا لواحد بين النقاط (x_1, x_2) و (y_1, y_2) بحيث يكون للمعادلتين $Y_2 = U_2(x_1, x_2)$ و $Y_1 = U_1(x_1, x_2)$ حلان وحيدان بالنسبة لـ x_1, x_2 هما $X_1 = \lambda_1(y_1, y_2)$ و $X_2 = \lambda_2(Y_1, Y_2)$ عندئذ يكون التوزيع المشترك للمتغيرين العشوائيين Y_1, Y_2 من الشكل :

$$h(y_1, y_2) = f[\lambda_1(y_1, y_2) , \lambda_2(y_1, y_2)] . |J|$$

حيث يمثل J جاكوبى الانتقال المحدد بالعلاقة :

$$J = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{vmatrix}$$

وحيث يمثل $\frac{\partial x_i}{\partial y_1}$ المشتق الجزئى لـ $\lambda i = w_1(y_1, y_2)$ بالنسبة لـ y_1 على اعتبار أن y_2 ثابتا وأن $i = 1,2$. وبنفس الطريقة نعرف بقية المشتقات الجزئية .

مثال (٥,٤)

بفرض أن x_1, x_2 يمثلان متغيرين عشوائيين مستمران بالتوزيع الاحتمالى المشترك :

$$f(x_1 , x_2) = \begin{cases} \frac{3}{32} x_1^2 \; x_2 & : 0 < x_1 < 2 \\ & \;\; 1 < x_2 < 3 \\ 0 & : \text{فيما عدا ذلك} \end{cases}$$

لنفتش عن التوزيع الاحتمالى المشترك للمتغيرين $y_2 = x_1 x_2$ و $y_1 = x_1^3$

الحل

لنحل المعادلتين $y_2 = x_1 . x_2$ و $y_1 = x_1^3$ بالنسبة لـ x_1, x_2 فنجد أن :

$$x_1 = y_1^{1/3} \; , \; x_2 = \frac{y_2}{y_1^{1/3}}$$

ومن هاتين المعادلتين نجد أن معين جاكوبى هو من الشكل :

$$\begin{vmatrix} \frac{1}{3y_1^{2/3}} & 0 \\ -\frac{y_2}{3y_1^{4/3}} & \frac{1}{y_1^{1/3}} \end{vmatrix} = \frac{1}{3y_1}$$

كما نلاحظ أن التقابلات تمثل واحداً لواحد بين مجموعة النقاط :

$$\{ (x_1, x_2) | 0 < x_1 < 2 \, , \, 1 < x_2 < 3 \}$$

ومجموعة النقاط :

$$\{ (y_1 , y_2) \mid \frac{y_2^3}{27} < y_1 < 8 \ , \ 0 < y_2 < 6 \}$$

من النظرية (٥,٤) نجد أن التوزيع الاحتمالى المشترك للمتغيرين Y_1, Y_2 هو من الشكل :

$$h (y_1 , y_2) = \frac{y_2}{32y_1}$$

أو

$$g (y_1 , y_2) = \begin{cases} \frac{y_2}{32y_1} & : \frac{y_2^3}{27} < y_1 < 8 , \ 0 < y_2 < 6 \\ 0 & : \text{فيما عدا ذلك} \end{cases}$$

تنحصر المشاكل الناشئة عن إيجاد توزيع الاحتمال للمتغير العشوائى $Y = U(X)$ عندما يكون X مستمراً فى الحالة التى لا يكون فيها التقابل واحداً لواحد بين مجموعة القيم x ومجموعة القيم y . ففى هذه الحالة نلاحظ أنه يقابل كل قيمة لـ y أكثر من قيمة واحدة لـ x . فمثلا إذا كان $f(x)$ موجبا فوق المجال $-1 < x < 2$ ومعدوما فيما عدا ذلك ، وإذا اعتبرنا أن التقابل هو من الشكل $Y = X^2$ ، فإننا نجد فى هذه الحالة أن $x = \pm \sqrt{y}$ من أجل $0 < y < 1$ وأن $x = \sqrt{y}$ من أجل $1 < y < 4$. وتوزيع الاحتمال لـ Y فوق المجال $1 < y < 4$ نوجده باستخدام النظرية (٥,٣)

$$g(y) = f[w(y)] \, | J | = \frac{f (\sqrt{y})}{2 \sqrt{y}} \ , \ 1 < y < 4$$

وفى الحالة $0 < y < 1$ يمكننا تجزئة المجال $-1 < x < 1$ لنحصل على الدالتين العكسيتين :

$$x = \begin{cases} -\sqrt{y} & : -1 < x < 0 \\ \sqrt{y} & \ \ 0 < x < 1 \end{cases}$$

وهكذا نجد أن كل قيمة لـ y يقابلها قيمة وحيدة لـ x فى كل مجال من مجالات التجزئة ، كما هو موضح على الشكل (٥,٣) . كما نلاحظ أن :

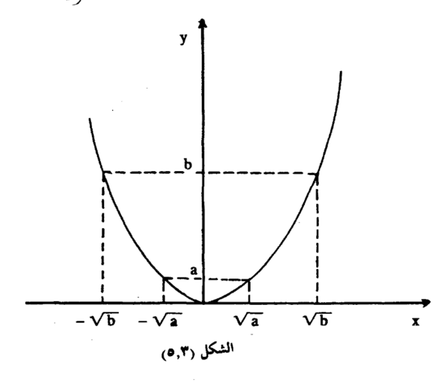
$$P[a < Y < b] = P[-\sqrt{b} < X < -\sqrt{a}] + P[\sqrt{a} < X < \sqrt{b}]$$

$$= \int_{-\sqrt{b}}^{-\sqrt{a}} f(x)\,dx + \int_{\sqrt{a}}^{\sqrt{b}} f(x)\,dx$$

وبتغير متحول التكامل من x إلى y نجد أن :

$$= \int_{b}^{a} f(-\sqrt{y})\, j_1 \cdot dy + \int_{a}^{b} f(\sqrt{y})\, J_2\, dy$$

$$= -\int_{a}^{b} f(-\sqrt{y})\, j_1 \cdot dy + \int_{a}^{b} f(\sqrt{y})\, J_2\, dy$$



الشكل (٥,٣)

حيث أن :

$$J_1 = \frac{d(-\sqrt{y})}{dy} = \frac{-1}{2\sqrt{y}} = -|J_1|$$

$$J_2 = \frac{d(\sqrt{y})}{dy} = \frac{1}{2\sqrt{y}} = |J_2|$$

ولذلك يمكن أن نكتب :

$$P[a < Y < b] = \int_a^b \left(f(-\sqrt{y})\,|J_1| + f(\sqrt{y})\,.\,|J_2| \right) dy$$

وهكذا نجد أن :

$$h(y) = f(-\sqrt{y})\,|J_1| + f(\sqrt{y})\,.\,|J_2|$$

$$= \frac{[f(-\sqrt{y}) + f(\sqrt{y})]}{2\sqrt{y}}, \quad 0 < y < 1$$

وأخيراً فإن :

$$h(y) = \begin{cases} \dfrac{[f(-\sqrt{y}) + f(\sqrt{y})]}{2\sqrt{y}}, & 0 < y < 1 \\ \dfrac{f(\sqrt{y})}{2\sqrt{y}}, & 1 < y < 4 \\ 0, & \text{فيما عدا ذلك} \end{cases}$$

يمكن تعميم الإجراءات السابقة فى البحث عن $h(y)$ من أجل $0 < y < 1$ بالنظرية (٥,٥) من أجل k دالة عكسية .

نظرية (٥,٥)

لنفرض أن X يمثل متغيراً عشوائياً مستمراً له توزيع احتمال $f(x)$ ، ولنفرض أن $Y = U(X)$ يمثل تقابلاً ليس واحداً لواحد بين قيم X وقيم Y . فإذا أمكن تجزئة مجال تحول المتغير X إلى k مجالاً جزئياً بحيث تمثل كل دالة عكسية من الدوال :

$$x_1 = \lambda_1(y)\ ,\ x_2 = \lambda_2(y)\ ,\ \ldots\ ,\ x_k = \lambda_k(y)$$

تقابلاً واحداً لواحد ، عندئذ يكون التوزيع الاحتمالى للمتغير Y من الشكل :

$$h(y) = \sum_{i=1}^{k} f[\lambda_i(y)]\,|J_i|$$

وحيث إن :

$$J_i = \lambda_i'(y) \quad , \quad i = 1, 2, \ldots, k$$

مثال (٥,٥)

إذا كان للمتغير X توزيعاً طبيعياً بالوسيطين μ ، σ ، فعندئذ يكون للمتغير توزيع كاى – مربع بدرجة حرية واحدة .

الحل

نلاحظ أن للمتغير $z = \frac{X-\mu}{\sigma}$ توزيعاً طبيعياً معيارياً وكثافة قدرها :

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$$

لنبحث الآن عن توزيع الاحتمال للمتغير $Y = z^2$. إن الحل العكسى للمعادلة $y = z^2$ هو $z = \pm \sqrt{y}$. لنرمز بـ $z_2 = +\sqrt{y}$ ، $z_1 = -\sqrt{y}$ نلاحظ أن

$$J_1 = -\frac{1}{2\sqrt{y}} \text{ و } J_2 = \frac{1}{2\sqrt{y}}$$

لذلك حسب النظرية (٥,٥) ولدينا :

$$g(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y} \cdot \left| \frac{-1}{2\sqrt{y}} \right| + \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y} \left| \frac{1}{2\sqrt{y}} \right|$$

$$= \frac{1}{2^{\frac{1}{2}}\sqrt{\pi}} e^{-\frac{y}{2}} \cdot y^{\frac{1}{2}-1} \quad , \quad y > 0$$

ومن المعلوم أن

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$$

لذلك فإن :

$$g(y) = \frac{1}{2^{\frac{1}{2}}\Gamma\left(\frac{1}{2}\right)} y^{\frac{1}{2}-1} \cdot e^{-\frac{1}{2}y} , \quad y > 0$$

وهذا يعنى أن للمتغير Y توزيعاً من نوع توزيع كاى – مربع بدرجة واحدة من الحرية .

(٥,٢) الدوال المولدة للعزوم Moments generating functions

على الرغم من أن طريقة تغير المتغيرات تقودنا إلى طريقة فعالة فى إيجاد توزيع دالة بعدة متغيرات ، فإن هناك أيضا إجراءات مفضلة فى الحالة التى تكون فيها المتغيرات مستقلة . وسوف نشير إلى هذه الإجراءات من خلال ما يسمى بالدوال المولدة للعزوم .

تعريف (٥,١) الدالة المولدة للعزوم

نعرف الدالة المولدة للعزوم لمتغير عشوائى X والتى نرمز لها بالرمز $M_x(t)$ بإحدى العلاقتين التاليتين .

$$M_x(t) = E(e^{t \cdot x})$$

$$= \sum_x e^{tx} \cdot f(x) \qquad : \qquad \text{إذا كان X منقطعاً}$$

$$= \int_{-\infty}^{+\infty} e^{tx} \cdot f(x)\, dx \qquad : \qquad \text{إذا كان X مستمراً}$$

وتكون الدالة المولدة للعزوم موجودة إذا كان المجموع أو التكامل فى التعريف (٥,١) محدوداً . فإذا كانت الدالة المولدة للعزوم موجودة ، فيمكن عندئذ استخدامها فى إيجاد مختلف عزوم المتغير العشوائى ، وطريقة ذلك توضحها النظرية التالية .

نظرية (٥,٦)

إذا كان $M_x(t)$ ممثلا للدالة المولدة للعزوم للمتغير العشوائى X فعندئذ يكون :

$$\left. \frac{d^r M_x(t)}{dt^r} \right|_{t=0} = \mu'_r$$

البرهان

من التعريف (٥,١) وباشتقاق ما تحت إشارة المجموع أو التكامل فإننا نحصل على ما يلى :

$$\frac{d^r M_x(x)}{dt^r} = \sum_{i=1}^{n} x_i^r \cdot e^{tx_i} f(x_i)$$

أو :

$$\frac{d^r M_x(x)}{dt^r} = \int_{-\infty}^{+\infty} x^r \cdot e^{tx} f(x)\, dx$$

وبوضع $t = 0$ فى كلا طرفى العلاقة السابقة فإننا نجد أن :

$$\mu_r' = E(X^r) = \left.\frac{d^r M_x(x)}{dt^r}\right|_{t=0}$$

مثال (٥,٦)

لنبحث عن الدالة المولدة للعزوم للمتغير العشوائى الحدانى X . من التعريف (٥,١) نجد أن :

$$M_x(t) = \sum_{x=0}^{n} e^{tx} \cdot \binom{n}{x} p^x \cdot q^{n-x}$$

$$= \sum_{x=0}^{n} \binom{n}{x} (p \cdot e^t)^x \cdot q^{n-x}$$

وهكذا نجد أن

$$M_x(t) = (p\, e^t + q)^n$$

وحسب النظرية (٥,٦) نجد بالاشتقاق مرتين أن :

$$\frac{dM_x(t)}{dt} = n \cdot p\, (p\, e^t + q)^{n-1} \cdot e^t$$

ومنه :

$$\mu = \mu_1' = \left.\frac{dM_x(t)}{dt}\right|_{t=0} = n \cdot p$$

وهو ما برهنا عليه سابقا . كذلك فإن :

$$\frac{d^2M_x(t)}{dt^2} = n \cdot p\,[\,e^t\,(p\,e^t + q)^{n-1} + p\,(n-1)\,(p\,e^t + q)\,e^{2t}\,]$$

وبوضع $t = 0$ نجد أن :

$$\mu_2 = n\,p\,[\,1 + p(n-1)]$$

وهكذا نجد أن :

$$\sigma^2 = \mu_2 - \mu^2 = np + np^2\,(n - 1) - n^2\,p^2$$

$$= np - np^2$$

$$= np\,(1 - p)$$

$$= np.\,q$$

وهى نفس النتيجة التى وجدناها فى الفصل الثالث .

مثال (٥,٧)

لنبين أن الدالة المولدة للعزوم لمتغير عشوائى طبيعى بالتوقيع μ والتباين σ^2 هى من الشكل :

$$M_x(t) = e^{\mu t + \frac{1}{2}\sigma^2 . t^2}$$

الحل

بالحقيقة : من التعريف (٥,١) نجد أن :

$$M_x(t) = \int_{-\infty}^{+\infty} e^{tx} . \frac{1}{\sqrt{2\pi}\,.\,\sigma}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

$$= \frac{1}{\sqrt{2\pi}\,\sigma} \int_{-\infty}^{+\infty} e^{-\frac{1}{2\sigma^2}[X^2 - 2tx\sigma^2 + \mu^2 - 2\mu X]}\, dx$$

وبإكمال المقدار $I = x^2 - 2(\mu + t\sigma^2)x + \mu^2$ إلى مربع كامل ، فإننا نجد أن :

$$I = [x - (\mu + t\sigma^2]^2 - 2\mu t\sigma^2 - t^2\sigma^4$$

وهكذا نجد أن :

$$M_x(t) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{\mu t + \frac{1}{2}t^2\sigma^2} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}\left[\frac{x-(\mu + t\sigma^2)}{\sigma}\right]^2} dx$$

وبإجراء تغيير فى المتحول بالعلاقة $4 = \frac{x - (\mu + t\sigma^2)}{\sigma}$ نجد أن $dx = \sigma du$

$$M_x(t) = e^{\mu t + \frac{1}{2}t^2\sigma^2} \cdot \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}u^2} du$$

$$= e^{\mu t + \frac{1}{2}t^2\sigma^2}$$

وهو المطلوب .

مثال (٥,٨)

بفرض أن للمتغير X توزيعاً من نوع كاى – مربع بـ ν درجة من الحرية ، فعندئذ تكون الدالة المولدة للعزوم لهذا المتغير من الشكل :

$$M_x(t) = \frac{1}{(1 - 2t)^{\nu/2}}$$

الحل

لقد حصلنا على توزيع كاى – مربع على شكل حالة خاصة من توزيع غما بوضع $\beta = 2, \alpha = \frac{\nu}{2}$. باستبدال قيمة f(x) فى التعريف (٥,١) نجد أن :

$$M_x(t) = \int_0^{\infty} e^{tx} \frac{1}{2^{\nu/2}\,\Gamma(\frac{\nu}{2})} x^{\frac{\nu}{2}-1} \cdot e^{-\frac{x}{2}} dx$$

$$= \frac{1}{2^{\nu/2}\,\Gamma(\frac{\nu}{2})} \int_0^{\infty} x^{\frac{\nu}{2}-1} \cdot e^{-x(\frac{1-2t}{2})} dx$$

وبوضع $y = \left(\frac{1-2t}{2}\right) x$ ومنه $dx = \frac{2}{1-2t} dy$:

$$M_x(t) = \frac{1}{2^{\nu/2}\,\Gamma(\frac{\nu}{2})} \int_0^\infty \left(\frac{2y}{1-2t}\right)^{\frac{\nu}{2}-1} \cdot e^{-y} \cdot \left(\frac{2}{(1-2t)}\right) dy$$

$$= \frac{1}{\Gamma(\frac{\nu}{2})\,(1-2t)^{\nu/2}} \int_0^\infty y^{\frac{\nu}{2}-1}\, e^{-y}\, dy$$

وإذا لاحظنا أن :

$$\int_0^\infty y^{\frac{\nu}{2}-1}\, e^{-y}\, dy = \Gamma\left(\frac{\nu}{2}\right)$$

لوجدنا أن :

$$M_x(x) = \frac{1}{(1-2t)^{\nu/2}}$$

لنستعرض الآن أربع نظريات تمثل خواص الدوال المولدة للعزوم . هذه الخواص ستكون مفيدة جداً فى إيجاد توزيع أى تركيب خطى لمجموعة متغيرات عشوائية مستقلة . أما النظرية الأولى (٥,٧) فسنوردها بدون برهان .

نظرية (٥,٧) (نظرية الوحدانية)

إذا كان للمتغيرين العشوائيين Y و X نفس الدالة المولدة للعزوم ، فعندئذ يكون لهما نفس توزيع الاحتمال .

نظرية (٥,٨)

$$M_{X+a}(t) = e^{at} \cdot M_X(t)$$

البرهان

$$M_{x+a}(t) = E[e^{t(X+a)}] = e^{at} E[e^{t}] = e^{at} M_x(t)$$

نظرية (٥,٩)

$$M_{aX}(t) = M(a.t)$$

البرهان

$$M_{aX}(t) = E[e^{t(aX)}] = E[e^{(ta)X}]$$

$$= M_X(a \cdot t)$$

نظرية (٥,١٠)

إذا كان المتغيران العشوائيان X_1, X_2 مستقلين ، وكان لهما دالتان مولدتان للعزوم $M_{X_1}(t)$ ، $M_{X_2}(t)$ على الترتيب ، فعندئذ يكون للمتغير $Y = X_1 + X_2$ دالة مولدة للعزوم من الشكل :

$$M_Y(t) = M_{X_1+X_2}(t) = M_{X_1}(t) \cdot M_{X_2}(t)$$

البرهان

$$M_Y(t) = E(e^{ty}) = E[e^{t(x_1 + x_2)}]$$

$$= \int\int_{-\infty}^{+\infty} e^{t(x_1 + x_2)} f(x_1, x_2)\, dx_1\, dx_2$$

وبما أن المتغيرين X_1, X_2 مستقلان فهذا يعني أن $f(x_1, x_2) = g(x_1)\, h(x_2)$ ولذلك فإن :

$$M_y(t) = \int_{-\infty}^{+\infty} g(x_1)\, e^{tx_1} dx_1 \int_{-\infty}^{+\infty} h(x_2) \cdot e^{tx_2} dx_2$$

$$= M_{X_1}(t) \cdot M_{X_2}(t)$$

ويتم برهان النظرية السابقة من أجل عدة متغيرات منقطعة بنفس الطريقة وذلك بأخذ المجموع بدلا عن التكامل .

باستخدام النظريتين (٥,٧) ، (٥,١٠) مع الانتباه إلى أن الدالة المولدة للعزوم لمتغير عشوائى بواسونى هى من الشكل $M_x(t) = \; = e^{\mu(e^t - 1)}$

نجد أنه إذا كان X_2, X_1 متغيرين من النوع البواسونى وكانا مستقلين ، وإذا كان الدالتان المولدتان للعزوم لهما على الترتيب $M_{X_1}(t) = e^{\mu_1 (e^t - 1)}$, $M_{X_2}(t) = e^{\mu_2(e^t - 1)}$ فعندئذ تكون الدالة المولدة للعزوم للمتغير $Y = X_1 + X_2$ من الشكل :

$$M_Y(t) = M_{X_1}(t) \,.\, M_{X_2}(t)$$

$$= e^{\mu_1(e^t - 1)} \,.\, e^{\mu_2(e^t - 1)}$$

$$= e^{(\mu_1 + \mu_2)(e^t - 1)}$$

وهذا يعنى أن مجموع متغيرين بواسونيين (بالتوقعين μ_1 , μ_2) هو متغير بواسونى وسيطه يساوى مجموع الوسيطين $\mu_1 + \mu_2$.

وغالبا ما نحتاج فى التطبيقات الإحصائية إلى تعريف التوزيع الاحتمالى لأى تركيب خطى بمتغيرات عشوائية طبيعية . لنفتش عن التوزيع الاحتمالى للمتغير $Y = a_1 X_1 + a_2 X_2$ إذا كان X_i طبيعيا بالتوقع μ_i والتباين σ_i حيث $i = 1,2$ ، وهذه المتغيرات مستقلة أيضا .

نلاحظ أولا حسب النظرية (٥,١٠) أن

$$M_Y(t) = M_{a_1 X_1}(t) \cdot M_{a_2 X_2}(t)$$

وباستخدام النظرية (٥,٩) نجد أن :

$$= M_{X_1}(a_1 t) \cdot M_{X_2}(a_2 t)$$

وباستخدام عبارة الدالة المولدة للعزوم لمتغير طبيعى والتى استخرجناها فى المثال (٥,٧) نجد أن :

$$= e^{a_1 t \mu_1 + a_1^2 \sigma_1^2 t^2/2} \cdot e^{a_2 t \mu_2 + a_2^2 \sigma_2^2 t^2/2}$$

$$= e^{t(a_1 \mu_1 + a_2 \mu_2) + \frac{1}{2} t^2 (a_1^2 \sigma_1^2 + a_2^2 \sigma_2^2)}$$

وبمقارنة العبارة الاخيرة بنتيجة المثال (٥,٧) نجد أن لـ Y توزيعاً طبيعياً بالوسيطين $(a_1 \mu_1 + a_2 \mu_2)$, $(a_1^2 \sigma_1^2 + a_2^2 \sigma_2^2)$ ولتعميم هذه الحالة على n متغيرا طبيعيا نسوق النظرية التالية بدون برهان .

نظرية (٥,١١)

إذا كانت المتغيرات $X_1, X_2, \ldots, X_n$ مستقلة وطبيعية بالتوقعات $\mu_1, \ldots, \mu_n$ والتباينات $\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2$ على الترتيب ، فعندئذ يكون للمتغير العشوائى :

$$Y = a_1 X_1 + a_2 X_2 + \ldots + a_n X_n$$

توزيعا طبيعيا بالتوقع :

$$\mu_y = a_1 \mu_1 + a_2 \mu_2 + \ldots + a_n \mu_n$$

والتباين :

$$\sigma_y^2 = a_1^2 \sigma_1^2 + \ldots + a_n^2 \sigma_n^2$$

وهكذا يبدو واضحاً الآن أن التوزيعين البواسونى ، والطبيعى يتمتعان بالخاصة التالية :

مجموع أى عدد من هذه المتغيرات المستقلة هو من جديد متغير عشوائى له نفس شكل التوزيع الذى تعود إليه هذه المتغيرات . وتنطبق هذه الخاصة أيضا على توزيع كاى – مربع .

نظرية (٥,١٢)

إذا كان للمتغيرات المستقلة $X_1, \ldots X_n$ توزيعاً من نوع توزيع كاى – مربع بعدد من درجات الحرية $\nu_1, \ldots, \nu_n$ على الترتيب . فعندئذ يكون للمتغير

$$Y = X_1 + \ldots + X_n$$

توزيعا من نوع كاى – مربع بعدد من درجات الحرية مساوٍ لـ $\nu = \nu_1 + \ldots \nu_n$

البرهان

بحسب النظرية (٥,١٠) نجد أن :

$$M_y(t) = M_{x_1}(t) \dots . M_{x_n}(t)$$

وحسب التمرين (٥,٨) لدينا :

$$M_{x_i}(t) = \frac{1}{(1 - 2t)^{\frac{\nu_i}{2_i}}} : \quad i = 1, 2, \dots, n$$

وبالتعويض نجد أن :

$$M_Y(t) = \frac{1}{(1-2t)^{\frac{\nu_1 + \cdots + \nu_n}{2}}}$$

نتيجة (٥,١)

إذا كانت المتغيرات المستقلة $X_1, \dots, X_n$ طبيعية بالتوقع μ ، والتباين σ^2 فعندئذ يكون للمتغير

$$Y = \sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma} \right)^2$$

توزيعا من نوع كاى – مربع بـ n = درجة من الحرية .

إن النتيجة السابقة تنتج مباشرة من التمرين (٥,٥) والنظرية (٥,١٢) .

(٣,٥) العينة العشوائية Random sampling

يمكن أن نعبر عن نتيجة تجربة إحصائية إما بقيمة عددية أو بتمثيل وصفى . فمثلا عند إلقاء حجرى نرد ، إذا كنا نهتم بمجموع العددين اللذين ظهرا فإن نتيجة التجربة تمثلها بقيمة عددية . ومع ذلك إذا أجريت فحوص دم لطلاب مدرسة معينة ، وإذا كنا

نهتم بنوع الدم عند كل طالب ، فعندئذ يكون التمثيل الوصفى أكثر فائدة فى هذه الحالة . فدم الإنسان يمكن تصنيفه بثمانية طرق . وهذه الطرق هى : A , B, O أو AB مع إشارة زائدة أو ناقص حسب وجود أو فقدان المولد المضاد Rh .

ويتعامل الإحصائيون قبل كل شيء مع الملاحظات العددية . ففى تجربة اختبار نوع الدم فإن الإحصائى يرمز للأعداد من 1 إلى 8 لنوع دم كل طالب مفحوص ، وبعد ذلك يسجل الأعداد المخصصة من أجل كل طالب ، ويجد عدداً من الملاحظات مساوياً لعدد الطلاب الموجودين فى المدرسة . وبهذا يكون لدى الإحصائى عددا منتهيا من النتائج .

وفى تجربة إلقاء حجرى نرد إذا كان اهتمام الإحصائى ينصب على مجموع الوجهين اللذين ظهرا ، وإذا ألقى هذين الحجرين عدداً لانهائيا من المرات فسيحصل على مجموعة غير منتهية من العناصر ، وكل ملاحظة ستمثل نتيجة إلقاء معين .

إن المجموع الكلى للملاحظات التى تهمنا ، سواء كان منتهيا أم غير منته يؤلف ما يدعى بالمجتمع (population) . فى السنوات الماضية كانت كلمة مجتمع تعنى جماعة من الناس ، أما الآن فيستخدم الإحصائى هذا المصطلح لدراسة أى شيء يُهتم به سواء كان جماعة من الناس أو من الحيوانات أو الأشياء .

تعريف (٥,٢) المجتمع

إن المجموع الكلى للملاحظات التى نهتم بها يدعى بالمجتمع وعدد الملاحظات فى مجتمع ما يُعرف لنا بما يسمى حجم المجتمع . فمثلا إذا كان المجتمع المدروس هو إنتاج مصنع للأدوات الكهربائية ، وإذا كان الإنتاج العام لهذا المصنع فى اليوم $(10)^4$ لمبة ، فعندئذ نقول بأن حجم المجتمع المدروس هو $(10)^4$. كذلك فإن عدد أوراق اللعب فى ورق اللعب هو مجتمع منته . وعلى هذا الأساس فإن هناك نوعين من المجتمعات مجتمعا منتهياً ومجتمع غير منته .

إنَّ كل ملاحظة فى المجتمع تمثل قيمة لمتغير عشوائى X له نفس توزيع المجتمع f(x) ، وعندما نتكلم عن مجتمع طبيعى ، أو حدانى ، أو بشكل عام مجتمع f(x) فإن

المقصود بذلك هو مجتمع ملاحظاته تمثل قيماً لمتغير عشوائى له التوزيع الطبيعى أو الحدانى أو التوزيع الاحتمالى $f(x)$. كذلك فإن توقع وتباين متغير عشوائى فى مجتمع $f(x)$ هما نفس توقع وتباين المجتمع الموافق $f(x)$.

والذى يهم الإحصائى هو البحث بطريق الاستنتاج عن الوسطاء المجهولة للمجتمع . فمثلا يحوى المجتمع الطبيعى متغيرين هما σ^2 و μ وقد يكون أحد هذين الوسيطين أو كلاهما مجهولا . ولذلك فإن غاية الإحصائى تقدير هذين الوسيطين من المعلومات التى تقدمها العينة المستخرجة من المجتمع بصورة عشوائية . وهذا ما يقودنا إلى نظرية العينات . وإذا كانت استنتاجاتنا مضبوطة (دقيقة) ، فيجب علينا أن نفهم العلاقة بين العينة ومجتمعها ، وعلى العينة أن تكون ممثلة للمجتمع المأخوذة منه ، وعليها أن تكون عينة عشوائية ، بمعنى أن الملاحظات التى تتألف منها يجب أن تكون مستقلة وعشوائية .

لنأخذ مثلاً عينة حجمها n من المجتمع المدروس $f(x)$ ، ولنعرف المتغير العشوائى , X_i $i = 1, 2, \ldots, n$ الممثل للقياس i أو قيمة الملاحظة المدرُوسة . عندئذ تؤلف المتغيرات العشوائية $X_1, X_2, \ldots, X_n$ عينة عشوائية فى المجتمع $f(x)$ بالقيم العددية $x_1, \ldots, x_2$ فيما إذا كانت القياسات مجراة فى تجربة مكررة n مرة تحت نفس الظروف ، وعلى هذا الأساس يمكن أن نفرض أن المتغيرات $X_1, \ldots, X_n$ مستقلة ، ولكل واحد منها نفس توزيع المجتمع $f(x)$ ، وهذا يعنى أن توزيعات هذه المتغيرات هى $f(x_1), \ldots, f(x_n)$ على الترتيب وتوزيعها المشترك هو :

$$f(x_1 , \ldots , x_n) = f(x_1) \ldots f(x_n)$$

تعريف (٥,٣) حجم عينة عشوائية Volume of random sample

إذا كان للمتغيرات المستقلة $X_1,\ldots, X_n$ نفس التوزيع الاحتمالى $f(x)$ ، عندئذ نقول بأن المتغيرات $X_1,\ldots, X_n$ تؤلف عينة عشوائية حجمها n من المجتمع $f(x)$ ، ونكتب توزيعها المشترك على النحو التالى .

$$f(x_1, \ldots , x_n) = f(x_1) \ . \ \ldots , \ f(x_n)$$

(٥,٤) نظرية العينات Sampling theory

إن هدفنا الرئيسى من سحب عينة عشوائية هو الحصول على معلومات حول الوسطاء المجهولة فى المجتمع المدروس . لنفرض أننا نريد الوصول عن طريق الاستنتاج لمعرفة نسبة الناس الذين يفضلون نوعا معينا من القهوة فى مدينة معينة . فمن المستحيل أن نسأل كل إنسان فى هذه المدينة عن رأيه وإحصاء المتغير الممثل للنسبة الحقيقية للأشخاص الذين يفضلون هذا النوع من القهوة . وعوضا عن ذلك ، فإننا نأخذ عينة عشوائية كبيرة من الناس ونسأل كل واحد منهم عن رأيه ، ونستنتج نسبة الذين يفضلون هذا النوع من القهوة . إن هذه القيمة سنستخدمها لاستنتاج بعض الحقائق المتعلقة بالنسبة الحقيقية فى هذه المدينة .

إن القيمة المحسوبة من خلال عينة تدعى إحصاء . وبما أنه يمكن أن نستخرج من مجتمع واحد عدة عينات عشوائية وبالتالى عدة إحصاءات ، لذلك علينا أن نتوقع بعض الاختلاف البسيط فى هذه الإحصاءات من عينة إلى أخرى . ولذلك فإن الإحصاء هو متغير عشوائى .

تعريف (٥,٤)

الإحصاء هو متغير عشوائى يعتمد فقط على قيم العينة العشوائية . نرمز عادة للإحصاء بأحد الأحرف اللاتينية الكبيرة . إن نسبة العينة فى المثال السابق هو إحصاء نرمز له عادة بالحرف $\hat{P}$. كما أن قيمة المتغير العشوائى $\hat{P}$ (من أجل عينة ما) والتى نرمز لها بالرمز $\hat{p}$ تستخدم كتقدير للنسبة الحقيقية $\hat{p}$ من الناس فى المدينة المفروضة والذين يفضلون هذا النوع من القهوة ، وعلينا أن نعرف التوزيع الاحتمالى للإحصاء $\hat{P}$.

سنعرف إحصاءَين هامين هما الوسط (the mean) والمتوسط (the median) . ويعتبر الوسط من أهم الإحصاءات لأنه يعبر عن تمركز التوزيع ، وهو أحد الإحصاءات المفيدة والأكثر استخداما ، وهو ما يشار إليه باسم الوسط الحسابى .

تعريف (٥,٥) وسط العينة Sample mean

لتكن $X_1, \ldots, X_n$ عينة عشوائية حجمها n . نُعَرِّف وسط العينة (sample mean) بالإحصاء كما يلي :

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

نلاحظ أن الاحصاء $\overline{X}$ يفترض القيمة $\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}$ (عندما يفترض X_1 القيمة x_1 ، X_2 القيمة x_2 ، وهلم جرا) .

مثال (٥,٩)

أوجد وسيط العينة الممثلة للملاحظات 15, 17, 19 .

الحل

إن القيمة الملاحظة للإحصاء $\overline{X}$ هى :

$$\overline{x} = \frac{15 + 17 + 19}{3} = 17$$

ونذكر هنا أن الإحصاء $\overline{X}$ سيُستخدم كتقدير لمتوسط المجتمع μ المجهول . والإحصاء المفيد الثانى ، والذى يقيس لنا مركز مجموعة من البيانات ، يدعى بالوسيط ، وسنرمز له بالرمز $\tilde{X}$.

تعريف (٥,٦) متوسط العينة Sample median

لتكن $X_1, \ldots, X_n$ عينة عشوائية حجمها n مرتبة تصاعدياً وفقاً لقيمها . نُعرِّف متوسط العينة (sample median) بأنه الإحصاء :

$$\tilde{x} = \begin{cases} \dfrac{x_{\frac{n}{2}} + x_{\left(\frac{n}{2}\right)+1}}{2} & \text{إذا كانت n زوجية :} \\ x_{\frac{n+1}{2}} & \text{إذا كانت n فردية :} \end{cases}$$

مثال (٥,١٠)

أوجد متوسط العينة العشوائية 7, 3, 4, 5, 5, 8, 7

الحل

بترتيب الملاحظات ترتيباً تصاعدياً وفقا لقيمها نجدها كالآتي : 3, 4, 5, 5, 7, 7, 8 .
والملاحظ أن $n = 7$ فردية لذلك فإن $\tilde{x} = x_{\frac{n+1}{2}}$ ومنه $\tilde{x} = x_4 = 5$ إذن فالوسط هو
$\tilde{x} = 5$

مثال (٥,١١)

أوجد متوسط عينة عشوائية ملاحظاتها 6 ,5 ,4 .9

الحل

لنرتب قبل كل شيء هذه الملاحظات ترتيباً تصاعدياً وفقالقيمهافنجدها كالآتي,4, 5, 6
9 . ونلاحظ أن $n = 4$ زوجية إذا

$$\tilde{x} = \frac{x_{\frac{4}{2}} + x_{\frac{4}{2}+1}}{2} = \frac{x_2 + x_3}{2}$$

ومنه $\tilde{x} = \frac{6+5}{2}$ وأخيرا فإن الوسيط هو $\tilde{x} = 5.5$

بعد أن حددنا تمركز توزيع المعلومات الإحصائية لنحاول الآن أن نقدم قياساً لتباين هذه المعلومات . لنأخذ مثلا مجموعتى الملاحظات الآتية :

3, 4, 5, 6, 8, 9, 10, 12, 15

3, 8, 8, 9, 9, 9, 10, 10, 15

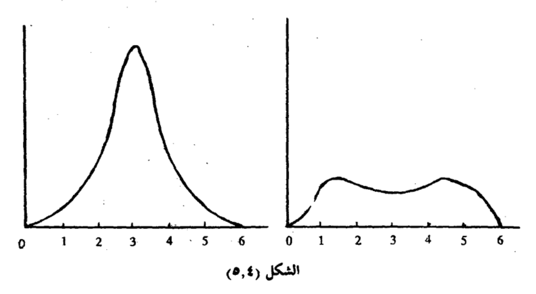
فنلاحظ أن لهما نفس المدى 12 (يعرف المدى بأنه الفرق بين أكبر وأصغر رقم) . كما نلاحظ فى مجموعة الملاحظات الأولى أن المتوسط والوسط متساويان وكلاهما يساوي 8 ، إلا أنه يوجد فرق كبير فى تباعد القياسات عن الوسط . فبالنسبة للمجموعة الثانية من الملاحظات ، فإننا نلاحظ أن المتوسط والوسط متساويات ويساوى كلا منهما العدد 9 ، ولكن غالبية الأعداد قريبة من الوسط . ولخاصة التغير هذه أهمية كبيرة فى المعلومات الإحصائية . فبالإضافة لأهمية العملية فهى تشكل قياسا ضروريا إلى جانب قياس النزعة المركزية (وهو القياس المعبر عن موضع تمركز التوزيع أى الوسط الحسابى مثلاً) لبناء صورة ذهنية لتوزيع التواتر . هذا وتوجد قياسات عديدة للتباين سنناقش هنا أهمها ، وسنبدأ بأبسطها وهو المدى .

تعريف (٥,٧) المدى The range

إن مدى عينة عشوائية $X_1, \ldots, X_n$ مرتبة تصاعدياً وفقاً لقيمها يتحدد بالإحصاء $X_n - X_1$.

مثال (٥,١٢)

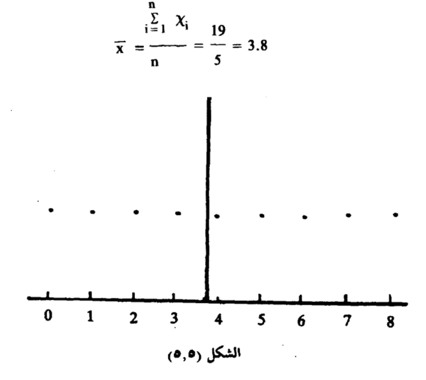
إن مدى الملاحظات 24 ,22 ,19 ,18 ,12 ,12 ,10 هو $24 - 10 = 14$. إذا كان حجم العينة كبيراً بقدر كاف ، فعندئذ يعتبر المدى قياساً رديئاً لتغيرات العينة ، لأنه لا يقدم لنا أى معلومات عن الطريقة التى تتوزع بها هذه القيم فيما بينها ، ولهذا السبب يعتبر المدى قياساً غير مرض تماما للتباين . لنتأمل على سبيل المثال التوزيعين الموضحين على الشكل (٥,٤) .



الشكل (٥,٤)

فنلاحظ أن لكل منهما نفس المدى إلا أن تباين الشكل ٤ ,٥ أ أكبر من تباين الشكل ٤ ,٥ ب . لنرى الآن فيما إذا كان بالإمكان إيجاد قياس للتباين يمكن التعبير عنه من خلال عدد ، وبحيث يكون أكثر حساسية من المدى . لذلك نأخذ على سبيل المثال مجموعة القياسات 5,7,1,2,4 . يمكن تمثيل هذه القياسات بيانياً كما فى الشكل (٥,٥) بوضع نقاط من أجل القياسات على طول محور الفواصل . ولحساب الوسط نجد أن :

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{19}{5} = 3.8$$



الشكل (٥,٥)

ونمثل الوسط بنقطة على محور الفواصل . ويمكن النظر إلى التباين الآن من خلال المسافة بين كل النقاط (القياسات) والوسط $\overline{X}$. فإذا كانت المسافة كبيرة قلنا أن المعلومات الإحصائية أكثر تباينا مما لو كانت المسافات أصغر . ونُعرِّف انحراف قياس X_i بأنه المسافة بين هذا القياس والوسط ($X_i - \overline{X}$) . ونلاحظ أن انحرافات القياسات الواقعة على يمين الوسط موجبة ، بينما انحرافات القياسات الواقعة على يسار الوسط سالبة . وفى المثال السابق يمكن تمثيل الانحرافات بالجدول (٥,١) .

X_i	$X_i - \overline{X}$	$(X_i - \overline{X})^2$	X_i^2
5	1.2	1.44	25
7	3.2	10.24	49
1	−2.8	7.84	1
2	−1.8	3.24	4
4	0.2	0.04	16
$\sum_{i=1}^{5} X_i = 19$	0	22.8	95

فإذا اتفقنا على أن الانحرافات تقدم معلومات عن التباين ، فإن من واجبنا الآن أن نضع علاقة واضحة عن هذه الانحرافات تزودنا بمقياس جيد للتباين .

وكخطوة جيدة فى هذا الطريق نختار ما يسمى بمربع الانحرافات .

تعريف (٥,٨) تباين العينة Sample variance

لتكن $X_1, \ldots, X_n$ عينة عشوائية حجمها n عندئذ نعرف تباين العينة بالإحصاء :

$$S^2 = \frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{n - 1}$$

لنرمز لقيمة S^2 من أجل عينة معينة بالرمز S^2 . والملاحظ أن S^2 تمثل وسط مربعات انحرافات الملاحظات عن أوساطها . أما سبب استخدام $(n - 1)$ عوضا عن n ، فسيظهر فى الفصل السادس .

نظرية (٥,١٣)

إذا كان S^2 ممثلاً لتباين عينة عشوائية ذات حجم n فعندئذ يمكننا أن نكتب S^2 على النحو التالى :

$$S^2 = \frac{n \sum_{i=1}^{n} X_i^2 - \left(\sum_{i=1}^{n} X_i\right)^2}{n(n-1)}$$

البرهان

من تعريف S^2 نستطيع أن نكتب :

$$S^2 = \frac{\sum^{n} (X_i^2 - 2\bar{X}X_i + \bar{X}^2)}{n-1}$$

$$= \frac{\sum_{i=1}^{n} X_i^2 - 2\bar{X}\sum^{n} X_i + n\bar{X}^2}{n-1}$$

وبتبديل $\bar{X} = \frac{\Sigma X_i}{n}$ ، وضرب كل من بسط ومقام الكسر السابق بالعدد n فإننا نجد أن :

$$S^2 = \frac{n \sum^{n} X_i^2 - (\sum^{n} X_i)^2}{n(n-1)}$$

لنرمز للانحراف المعيارى العينى بالرمز S ، عندئذ يمثل S الجذر الموجب للتباين العينى .

مثال (٥,١٣)

بين أن التباين العينى للملاحظات 6, 9, 8, 7, 5, 6 هو $S^2 = 65$

الحل

نلاحظ أن :

$$\sum_{i=1}^{6} X_i^2 = 291 \quad , \quad \sum_{i=1}^{6} X_i = 41 \quad , \quad n=6$$

وحسب النظرية (٥,١٣) نجد أن :

$$S^2 = \frac{(6)(291) - (41)^2}{(6)(5)} = 65$$

مثال (٥,١٤)

احسب $S, \bar{X}$ لمجموعة القياسات 85 ,70 ,60 ,90 ,81

X_i	X_i^2
85	7225
70	4900
60	3600
90	8100
81	6561
386	30386

الحل

نلاحظ أن $\bar{X} = \frac{386}{5} = 77.5$ كما نلاحظ أن :

$$\sum_{i=1}^{n} (X_i - \bar{X})^2 = \Sigma X_i^2 - \frac{(\sum^{n} X_i)^2}{n}$$

ولذلك فإن :

$$S = \sqrt{\frac{\sum^{n} (X_i - \bar{X})^2}{n-1}} = \sqrt{\frac{586.8}{4}} = \sqrt{146.7} = 12.1$$

يشكل حقل الإحصاء الاستقرائى قاعدة تتعلق بما يسمى بالتعميم والتنبؤ ، هذا ويمكن إجراء التعميم من الإحصاء إلى المتغير بثقة . إذا فهمنا تغير سلوك هذا الإحصاء عند حسابه من خلال عينات مختلفة مأخوذة من نفس المجتمع المدروس . وسيعتمد توزع الإحصاء على حجم المجتمع ، وحجم العينة ، والطريقة التى تمت بها اختيار العينة الإحصائية . فإذا كان حجم المجتمع المدروس كبيراً جداً أو غير منته ، فعندئذ سيكون للإحصاء نفس التوزيع سواء سحبنا عينتنا مع الإعادة أو بدون إعادة . أما إذا احتوى مجتمعنا المدروس على عدد بسيط منته من العناصر ، فإن السحب مع الإعادة من هذا المجتمع سيقودنا إلى توزيع يختلف بشكل تافه عما إذا كنا سحبنا بدون إعادة . أن السحب مع الإعادة من مجتمع منته يكافىء سحباً من مجتمع غير منته . لذلك فإنه ليس هناك تحديد للحجم الممكن للعينة المسحوبة .

تعريف (٥,١٠) التوزيع الاحتمالى Probabilistics distribution

إن التوزيع الاحتمالى للإحصاء يسمى بتوزيع المعاينة .

تعريف (٥,١١) الانحراف المعيارى Standard deviation

يدعى الانحراف المعيارى لتوزيع المعاينة لإحصاء ما بالخطأ المعيارى لهذا الإحصاء .

ملاحظة

إن التوزيع الاحتمالى للإحصاء $\bar{X}$ يسمى بتوزيع المعاينة للوسط الحسابى ، والخطأ المعيارى للوسط هو الانحراف المعيارى لتوزيع المعاينة لـ $\bar{X}$. إن كل عينة ذات حجم n مأخوذة من مجتمع معين ستقودنا إلى قيمة s للإحصاء S ، وتسمى هذه القيمة بالانحراف المعيارى لتوزيع المعاينة . وعلى هذا الأساس فإن الخطأ المعيارى للانحراف المعيارى العينى ما هو إلا الانحراف المعيارى للإحصاء S .

فيما تبقى من هذا الفصل سنورد بعض توزيعات المعاينة الهامة والتى يستخدمها الإحصائى بشكل تكرارى . أما تطبيقات هذه التوزيعات ، فسنوردها فى الفصلين السادس والسابع .

(٥,٥) توزيع المعاينة للوسط

Sampling distribution of mean

إن أول وأهم توزيعات المعاينة هو توزيع الوسط $\bar{X}$. لنفرض ، على سبيل المثال ، عينة عشوائية ذات حجم n مأخوذة من مجتمع طبيعى بالمتوسط μ والتباين σ^2 . إن كل ملاحظة $X_i , i = 1, n$ من العينة العشوائية سيكون لها نفس التوزيع الطبيعى . لذلك حسب تعريف الوسط :

$$\bar{X} = \frac{X_1 + \dots + X_n}{n}$$

وحسب النظرية (٥,١١) نجد أن لـ $\bar{X}$ توزيعا طبيعيا بالتوقع :

$$\bar{\mu} = \frac{\mu_1 + \mu_2 + \dots + \mu_n}{n} = \mu$$

والتباين :

$$\sigma^2_{\bar{x}} = \frac{\sigma_1^2 + \sigma_2^2 + \dots + \sigma_n^2}{n^2} = \frac{\sigma^2}{n}$$

إذا سحبنا عينة من مجتمع توزيعه مجهول منته أو غير منته ، فعندئذ سيكون توزيع المعاينة للإحصاء $\bar{X}$ قريباً جداً من التوزيع الطبيعى بالتوقع μ والتباين $\frac{\sigma^2}{n}$ شريطة أن يكون حجم العينة n كبيراً . إن هذه النتيجة المذهلة تنتج مباشرة من النظرية التالية والتى نسميها عادة بنظرية النهايات المركزية (central limit theorem) والتى نؤجل برهانها فى هذا الكتاب .

نظرية (٥,١٤)

ليكن $\overline{X}$ وسط عينة عشوائية حجمها n مأخوذة من مجتمع إحصائى بالتوقع μ والتباين σ^2 . عندئذ يسعى التوزيع الاحتمالى للمتغير :

$$Z = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}}$$

إلى التوزيع الطبيعى المعيارى عندما تسعى n إلى ∞ .

إن التقريب الطبيعى للإحصاء $\overline{X}$ سيكون جيداً إذا كان $n \geq 30$ بغض النظر عن شكل المجتمع . أما إذا كان $n < 30$ فعندئذ سيكون التقريب جيدا فيما إذا كان المجتمع المدروس لا يختلف كثيراً عن المجتمع الطبيعى . أما إذا علم أن المجتمع طبيعياً ، فعندئذ سيكون للإحصاء توزيع معاينة طبيعى مهما كان حجم العينة n .

مثال (٥,١٥)

تنتج شركة للمصابيح الكهربائية نوعاً من اللمبات عمرها (مدة عملها) له تقريبا توزيعا طبيعيا بالوسط $\mu = 750$ ساعة ، والانحراف المعيارى $\sigma = 38$ ساعة . أوجد احتمال أن يكون لعينة مؤلفة من 14 لمبة وسطاً أقل من 735 ساعة .

الحل

نلاحظ أن توزيع المعاينة لوسط العينة X سيكون قريبا من الطبيعى بـ $\mu_{\bar{x}} = 750$ ، $\sigma_{\bar{x}} = \frac{38}{\sqrt{14}} = 10.155$. كما نلاحظ أن الاحتمال المطلوب يمثل مساحة القسم المظلل على الشكل (٥,٦) .

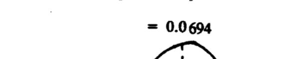
واعتماداً على كون $\overline{X}_1 = 735$ فإننا نجد أن :

$$Z_1 = \frac{735 - 750}{10.155} = -1.476$$

ولذلك فإن :

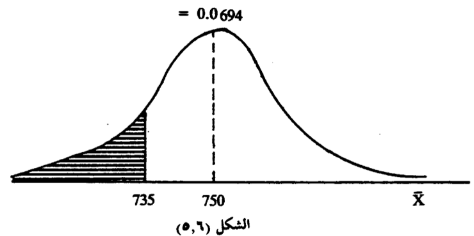
$$P[\bar{X} < 735] = P[Z < -1.48]$$

وحسب الجدول IV نجد أن :



$$P[\bar{X} < 735] = 0.0694$$

$$= 0.0694$$



الشكل (٥,٦)

مثال (٥,١٦)

بفرض أن المجتمع الإحصائى المدروس هو المجتمع المنقطع المنتظم بالكثافة :

$$f(x) = \begin{cases} \frac{1}{6}, & : x = 1, 2, 3, 4, 5, 6 \\ 0, & : \text{فيما عدا ذلك} \end{cases}$$

أوجد احتمال أن نسحب عينة حجمها n من هذا المجتمع مع الإعادة . فنجد أن وسط العينة أقل من 3.7 وأكبر من 3.2 إذا قيس الوسط بالنسبة لأقرب جزء عشرى .

الحل

لحساب وسط وتباين التوزيع المنتظم السابق فإننا نجد أن :

$$\mu_x = \frac{1 + 2 + 3 + 4 + 5 + 6}{6} = 3.5$$

$$\sigma_x^2 = \frac{(1-3.5)^2 + (2-3.5)^2 + \dots + (6-3.5)^2}{6} = 2.91$$

هذا ويمكن تقريب توزيع المعاينة لـ $\bar{X}$ إلى التوزيع الطبيعي بالوسط $\mu = 3.5$ والتباين $0.485 = \frac{\sigma^2}{n} = \sigma_{\bar{x}}^2$ ، وبأخذ الجذر التربيعي نجد أن $\sigma = 0.696$.

أما احتمال أن يكون $\bar{X}$ أقل من 3.7 وأكبر من 3.2 فيمثل بالقسم المظلل من الشكل (٥,٧) ، وقيم z الموافقة لـ $\bar{X}$ $\bar{X}_1 = 3.15$ و $\bar{X}_2 = 3.75$ فهي على الترتيب .

$$Z_1 = \frac{3.15 - 3.5}{0.696} = 0.502$$

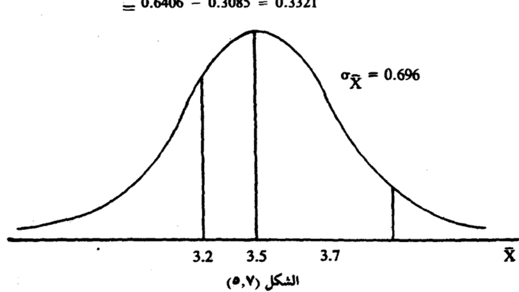
$$Z_2 = \frac{3.75 - 3.5}{0.696} = 0.359$$

لذلك فإن :

$$P[3.2 < \bar{X} < 3.7] \simeq p[-0.502 < Z < 0.359]$$

$$= P[Z < 0.359] - P[Z < -0.502]$$

$$\simeq 0.6406 - 0.3085 = 0.3321$$



الشكل (٥,٧)

لنفرض أن لدينا مجتمعين ، الأول بالوسط μ_1 والتباين σ_1^2 أما الثانى بـ $\bar{X}_1$ فبالوسط μ_2 والتباين σ_2^2 . لنرمز بـ $\bar{X}_2$ لوسط عينة حجمها n_1 مسحوبة بشكل عشوائى من المجتمع الأول وبـ $\bar{X}_2$ لوسط عينة حجمها n_2 مسحوبة من المجتمع الثاني وبشكل عشوائى أيضا . وبصورة مستقلة عن العينة الأولى . والسؤال المطروح الآن : كيف يتوزع الفرق $\bar{X}_2 - \bar{X}_1$ عند تكرار السحب ؟

للإجابة على هذا السؤال نلاحظ حسب النظرية (٥,١٤) أن للمتغيرين $\bar{X}_2$ ، $\bar{X}_1$ توزيعين قريبين من التوزيع الطبيعي بالوسطين μ_2 ، μ_1 والتباينين $\frac{\sigma_2^2}{n_2}$, $\frac{\sigma_1^2}{n_1}$ على الترتيب . ويتحسن هذا التقريب بازدياد كل من n_1 , n_2 وباختبار عينات مستقلة من المجتمعين نجد أن المتغيرين : $\bar{X}_2 , \bar{X}_1$ سيكونان مستقلين وعندئذ بحسب النظرية (٥,١١) وبفرض أن $q_1 = 1$ و $q_2 = -1$ نستنتج أن للمتغير $\bar{X}_1 - \bar{X}_2$ توزيعاً قريبا من التوزيع الطبيعي بالوسط :

$$\mu_{\bar{X}_1 - \bar{X}_2} = \mu_{\bar{X}_1} - \mu_{\bar{X}_2} = \mu_1 - \mu_2$$

والتباين :

$$\sigma^2_{\bar{X}_1 - \bar{X}_2} = \sigma^2_{\bar{X}_1} + \sigma^2_{\bar{X}_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

نظرية (٥,١٥)

لدى سحب عينات مستقلة حجمها n_1 , n_2 من مجتمعين منقطعين أو مستمرين بالـــوسطين μ_1 , μ_2 والتبايـــنين σ_1^2 , σ_2^2 على الترتـــيب ، فعندئـــذ سيكـــون للفـــرق $\bar{X}_1 - \bar{X}_2$ توزيعا قريبا من التوزيع الطبيعى بالوسط $\mu_{\bar{X}_1 - \bar{X}_2} = \mu_1 - \mu_2$

والتباين :

$$\sigma^2_{\bar{X}_1 - \bar{X}_2} = \frac{\sigma_1^2}{n_1} = \frac{\sigma_2^2}{n_2}$$

وحسب نظرية النهايات المركزية سيكون للمتغير :

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

توزيعاً قريباً من التوزيع الطبيعى المعيارى .

إذا كان كلا من n_1 , n_2 أكبر من 30 ، فعندئذ سيكون التقريب الطبيعى لتوزيع $\bar{X}_1 - \bar{X}_2$ جيداً جداً .

مثال (٥,١٧)

بفرض أن لقنوات التصوير التلفزيونية العائدة للمنتج A وسط عمر قدره ست سنوات ونصف وانحراف معيارى قدره 0.9 سنة . بينما لها وسط عمر قدره ست سنوات وانحراف معيارى قدره 0.8 سنة بالنسبة لمصنع يعود إلى المنتج B . ما هو احتمال أن يكون لعينة عشوائية مؤلفة من 36 قناة (عائدة للمنتج A) على الأقل وسط عمر أكبر بعام من وسط عمر عينة عشوائية مؤلفة من 49 قناة عائدة للمنتج B ؟

الحل

لنرتب بعض المعلومات السابقة فى الجدول التالى :

المجتمع الأول	المجتمع الثانى
$\mu_1 = 6.5$ $\sigma_1 = 0.9$ $n_1 = 36$	$\mu_2 = 6.0$ $\sigma_2 = 0.8$ $n_2 = 49$

باستخدام النظرية (٥,١٥) نلاحظ أن لتوزيع المعاينة للفرق $\bar{X}_1 - \bar{X}_2$ وسطاً وانحرافاً معيارياً قدرهما :

$$\sigma_{\bar{X}_1 - \bar{X}_2} = 6.5 - 6.0 = 0.5$$

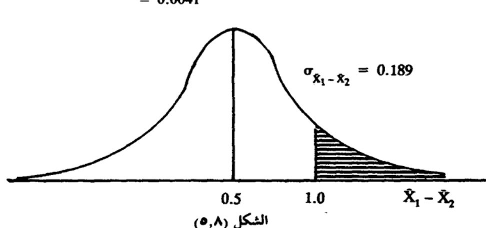
$$\sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{0.81}{36} + \frac{0.64}{49}} = 0.189$$

نلاحظ أن احتمال أن يكون لـ 36 قناة العائدة للمنتج A وسطا على الأقل أكبر بعام من وسط الـ 49 قناة العائدة للمنتج B يمكن تمثيله بالقسم المظلل على الشكل الموضح (٥,٨) . ويوافق القيمة $\bar{X}_1 - \bar{X}_2 = 1.0$ القيمة :

$$Z = \frac{1.0. - 0.5}{0.189} = 2.646$$

لذلك فإن :

$$P[\bar{X}_1 - \bar{X}_2 \geq 1.0] = P[Z > 2.646]$$
$$= 1 - P[Z < 2.646]$$
$$= 1 - 0.9959$$
$$= 0.0041$$



الشكل (٥,٨)

(٥,٦) توزيع المعاينة للمغير العشوائي $\frac{(n-1)S^2}{\sigma^2}$

إذا سحبنا عينة عشوائية ذات حجم n من مجتمع طبيعي بالوسط μ والتباين σ^2 ، وحسبنا تباين العينة S^2 ، فإننا نحصل على قيمة للإحصاء S^2 . وللإحصاء S^2 تطبيقات

عملية قليلة . وعوضاً عن هذا الإحصاء سندرس توزيع المتغير العشوائى $\frac{(n-1)\,S^2}{\sigma^2}$

بجمع وطرح العينة $\bar{X}$ فى العبارة $\sum_{i=1}^{n} (X_i - \mu)^2$ فإننا نجد أن :

$$\sum_{i=1}^{n} (X_i - \mu)^2 = \sum_{i=1}^{n} [(\bar{X}_i - \bar{X}) + (\bar{X} - \mu)]^2$$

$$= \sum_{i=1}^{n} (X_i - \bar{X})^2 + \sum_{i=1}^{n} (\bar{X} - \mu)^2 + 2(\bar{X} - \mu) \cdot \sum_{i=1}^{n} (X_i - \bar{X})$$

$$= \sum_{i=1}^{n} (X_i - \bar{X})^2 + n(\bar{X} - \mu)^2$$

وبتقسيم طرفى العلاقة السابقة على σ^2 وبوضع $(n-1)S^2$ بدلاً عن $\sum_{i=1}^{n} (X_i - \bar{X})^2$ فإننا نجد أن :

$$\frac{\sum^{n} (X_i - \mu)^2}{\sigma^2} = \frac{(n-1)\,S^2}{\sigma^2} + \frac{(\bar{X} - \mu)^2}{\sigma^2/n}$$

وحسب النتيجة (٥,١) نعلم أن للمتغير $\frac{\sum^{n}(X_i - \mu)^2}{\sigma^2}$ توزيعاً من نوع كاى – مربع بـ n درجة من الحرية . أما المتغير $\frac{(\bar{X} - \mu)^2}{\sigma^2/n}$ فله توزيع كاى – مربع بدرجة حرية واحدة وذلك حسب التمرين (٥,٥) وبما أن فرق متغيرين مستقلين لهما توزيع كاى – مربع هو من جديد متغير من نوع كاى مربع وبذلك يكون للمتغير $\frac{\sum^{n}(x_i - \bar{x})^2}{\sigma^2}$ توزيعا من نوع كاى – مربع بعدد درجات الحرية مساو لـ $(n-1)$.

نظرية (٥,١٦)

إذا كان S^2 ممثلا لتباين عينة عشوائية حجمها n مأخوذة من مجتمع طبيعى بالتباين σ^2 ، فعندئذ سيكون للمتغير العشوائى :

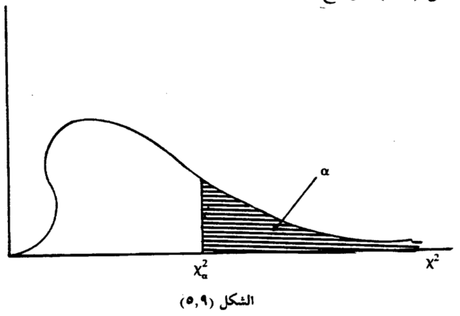
$$\chi^2 = \frac{(n-1)\,S^2}{\sigma^2}$$

توزيعا من نوع كاى – مربع بـ $\nu = n - 1$ درجة من الحرية .

إن قيم المتغير العشوائى x تحسب بالنسبة لكل عينة من العلاقة :

$$\chi^2 = \frac{(n-1)\,S^2}{\sigma^2}$$

كما أن احتمال أن تقدم عينة عشوائية قيمة χ^2 أكبر من قيمة مفروضة يساوى إلى المساحة الواقعة تحت المنحنى والواقعة إلى يمين المفروضة ، ومن المعتاد أن نرمز بـ χ^2_α لقيمة χ^2 التى تحدد لنا مساحة على يمينها قدرها α . وهذه المساحة يوضحها القسم المظلل على الشكل (٥,٩) الموضح أدناه .



الشكل (٥,٩)

الجدول السادس يوضح القيم المختلفة لـ χ^2_α الموافقة للقيم المختلفة لكل من α , ν . ونلاحظ فى هذا الجدول أن العمود الأيسر يوافق قيم ν المختلفة من 1 إلى 30 ، أما السطر العلوى فيعطينا المساحات α. لذلك فإن قيمة ^{2} (من أجل $\nu = 10$) والتى تحدد مساحة على يمينها قدرها 0.01 هى $\chi^2_{0.01} = 23.209$. وبسبب عدم وجود تناظر فى الشكل ، علينا أن نبحث أيضا عن $\chi^2_{0.99} = 2.558$ من أجل $\nu = 10$.

أن 99% من توزيع كاى – مربع بـ $n - 1$ درجة من الحرية يقع بين النقطتين $\chi^2_{0.995}$ $\chi^2_{0.005}$ ويشير إلى أن قيمة χ^2 الواقعة على يمين $\chi^2_{0.995}$ ليست مرجحة للوقوع إلا إذا افترضنا قيمة لـ χ^2 صغيرة جدا . وبشكل مشابه فإن قيمة χ^2 الواقعة على يسار ليست مرجحة للوقوع إلا إذا افترضنا قيمة لـ σ^2 كبيرة جدا . وبعبارة أخرى

يمكن أن يكون لـ X^2 قيما مرافقة على يسار القيمة $\chi^2_{0.995}$ أو على يمين القيمة $\chi^2_{0.005}$ إذا كانت σ^2 مضبوطة . وإذا صدف وحدث هذا فعلى الغالب نكون قد ارتكبنا خطأ فى قيمة σ^2 المفروضة .

مثال (٥,١٨)

تكفل شركة صناعية لإنتاج بطاريات السيارات بالعمل فى المتوسط أربع سنوات وبالانحراف المعيارى سنة . فإذا جربت أربع بطاريات ووجد أن عمر كل منها على النحو التالى 4.0 ,3.8 ,2.8 ,2.3 سنة ، فهل ستظل الشركة مقتنعة من أن لبطاريتها انحراف معيارى قدره عام ؟

الحل

لنبحث أولا عن تباين العينة :

$$S^2 = \frac{(4)\ (43.57)\ -\ (12.9)^2}{(4)\ (3)} = 0.6558$$

كما أن :

$$\chi^2 = \frac{(3)\quad(0.6558)}{1} = 1.9675$$

تمثل قيمة كاى – مربع بثلاث درجات من الحرية . لذلك فإن 99% من قيم χ^2 بثلاث درجات من الحرية ستقع بين النقطتين 12.838, 0.0717 والمحسوبتين من أجل $\sigma = 1$ وهى معقولة ، لذلك لا يوجد أى سبب لدى الشركة المنتجة للشك فى أن الانحراف المعيارى يمثل قيمة أخرى غير العام .

(٥,٧) التوزيع Distribution - t

قد لانكون محظوظين فى كثير من الأحيان فى معرفة تباين المجتمع الذى نختار منه العينات العشوائية . فمن أجل عينات حجمها $n \geq 30$ يكون S^2 تقديراً جيداً لتباين

المجتمع σ^2 والسؤال المطروح الآن : هل يحدث تغيرا فى توزيع الإحصاء $\frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$ فى النظرية (٥,١٤) إذا بدلنا σ^2 بـ S^2 ! بما أن S^2 تزودنا بتقدير جيدا ؟ لـ σ^2 ولا تتغير من عينة لأخرى من أجل $n \geq 30$ فإن توزيع الإحصاء $\frac{\bar{X} - \mu}{S / \sqrt{n}}$ يظل قريبا من التوزيع الطبيعى المعيارى .

أما إذا كان حجم العينة المختارة $(n < 30)$ فإن قيم S^2 ستتغير من عينة لأخرى وتوزيع المتغير العشوائى $\frac{\bar{X} - \mu}{S / \sqrt{n}}$ لن يكون قريباً من التوزيع الطبيعى المعيارى . سنذكر الآن متغيراً عشوائياً جديدا نرمز له بالرمز T ونعرفه على النحو التالى :

$$T = \frac{\bar{X} - \mu}{S / \sqrt{n}}$$

ولإيجاد توزيع المعاينة للمتغير T ، فإننا سنفترض أن العينة مسحوبة من مجتمع طبيعى ، وعندئذ نستطيع أن نكتب :

$$T = \frac{(\bar{X} - \mu) / (\sigma / \sqrt{n})}{\sqrt{S^2 / \sigma^2}} = \frac{Z}{\sqrt{v / (n - 1)}}$$

حيث يكون للمتغير :

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$$

توزيعاً طبيعياً معيارياً وللمتغير :

$$V = \frac{(n - 1) S^2}{\sigma^2}$$

توزيعاً من نوع كاى – مربع بـ $\nu = n - 1$ درجة من الحرية . ومن أجل عينات طبيعية ، يمكن أن نبين أن S^2 و $\bar{X}$ يمثلان متغيرين مستقلين ، وهكذا فإن V, Z يمثلان متغيرين مستقلين أيضا .

نظرية (٥,١٧)

ليكن Z متغيراً عشوائياً طبيعياً معيارياً ، V متغير من نوع كاى – مربع بـ درجة من الحرية . إذا كان المتغيران Z و V مستقلين ، فعندئذ يكون توزيع المتغير :

$$T = \frac{Z}{\sqrt{V/\nu}}$$

معرفة بالعلاقة :

$$h(t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\cdot\sqrt{\nu\cdot\pi}}\cdot\left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}} \quad : -\infty < t < +\infty$$

وهذا ما يعرف باسم التوزيع t بـ درجة من الحرية .

البرهان

بما أن المتغيرين العشوائيين V و Z مستقلان ، لذلك فإن التوزيع الاحتمالي المشترك لهما يعطى بالعلاقة :

$$f(Z, V) = \begin{cases} \frac{1}{\sqrt{2\pi}} e^{-\frac{Z^2}{2}} \cdot \frac{1}{2^{\nu/2}\,\Gamma\left(\frac{\nu}{2}\right)} e^{-\frac{1}{2}V}\, V^{\frac{\nu}{2}-1} & : -\infty < Z < +\infty \\ & \;\; 0 < V < \infty \\ 0 & : \text{فيما عدا ذلك} \end{cases}$$

لنعرف متغيراً جديداً $U = V$. إن الحل العكسى للمعادلتين $t = \frac{Z}{\sqrt{V/\nu}}$ و $u = V$ هما $V = u$ ، $Z = \frac{t\sqrt{u}}{\sqrt{\nu}}$ ، ونحصل منهما على :

$$J = \begin{bmatrix} \frac{\sqrt{u}}{\sqrt{\nu}} & \frac{t}{2\sqrt{u\cdot\nu}} \\ 0 & 1 \end{bmatrix} = \frac{\sqrt{u}}{\sqrt{\nu}}$$

ويمثل التقابل السابق تقابلاً واحداً لواحد بين عناصر المجموعة :

$$\{(Z, V) \mid -\infty < Z < +\infty, 0 < V < \infty\}$$

وعناصر المجموعة $\{(t, v) \mid -\infty < t < +\infty, 0 < U < \infty\}$ ، وباستخدام النظرية (٥,٤) نجد أن التوزيع الاحتمالي المشترك للمتغيرين u، t هو من الشكل :

$$g(t, u) = \begin{cases} \dfrac{1}{\sqrt{2\pi}\, 2^{\nu/2}\, \Gamma(\frac{\nu}{2})}\, e^{\frac{\nu}{2}-1} \cdot e^{-(\frac{\nu}{2})(1+\frac{t^2}{\nu})} \cdot \dfrac{\sqrt{u}}{\sqrt{\nu}} & : -\infty < t < +\infty,\ 0 < u < \infty \\ 0 & : \text{فيما عدا ذلك} \end{cases}$$

وبالمكاملة بالنسبة لـ u نجد أن التوزيع الاحتمالى للمتغير T هو من الشكل :

$$h(t) = \int_0^\infty g(t, u)\, du = \int_0^\infty \frac{1}{\sqrt{2\pi}\, 2^{\nu/2}\, \Gamma(\frac{\nu}{2})} u^{\frac{\nu}{2}-1} \cdot e^{-\frac{\nu}{2}\left(1+\frac{t^2}{\nu}\right)} \frac{\sqrt{u}}{\sqrt{\nu}}\, du$$

لنجرى الآن تغيراً فى المتحول بالعلاقة $Z = \dfrac{u\left(1+\frac{t^2}{\nu}\right)}{2}$ فنجد أن :

وهكذا فإن : $dv = dz \Big/ \left(1 + \dfrac{t^2}{v}\right)$

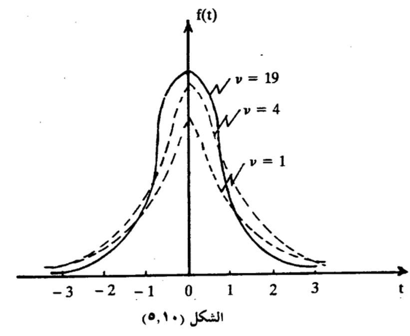
$$h(t) = \frac{1}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\pi\nu}} \left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}} \cdot \int_0^\infty Z^{\frac{\nu+1}{2}-1} \cdot e^{-z}\, dz$$

$$= \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right) \cdot \sqrt{\pi\nu}} \cdot \left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}} \quad : -\infty < t < +\infty$$

فيما عدا ذلك :

$$= \quad 0$$

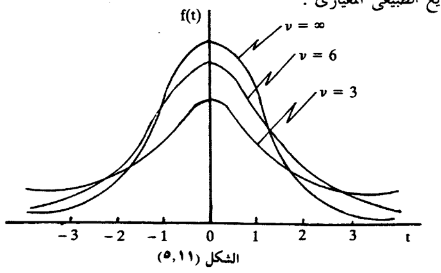
تدعى الدالة الأخيرة بدالة كثافة التوزيع t بـ ν درجة من الحرية حيث يمثل عدداً صحيحاً موجباً . ويقدم الشكل (٥,١٠) عدة تمثيلات بيانية لهذا المنحنى من أجل قيم متعددة لـ ν .



الشكل (٥,١٠)

نلاحظ أن التوزيع T متماثل بشكل مشابه للتوزيع Z . حيث أنهما متناظرين حول الوسط الذى يساوى الصفر . كما أن كلا التوزيعين له شكل المنحنى الجرسى ، غير أن التوزيع t أكثر تغيراً ، بسبب أن قيم T تعتمد على تغييرات قيم الكميتين S^2 و $\overline{X}$ فى حين أن قيم Z تعتمد على تغير $\overline{X}$ من عينة لأخرى . والملاحظ أيضا أن توزيع T يختلف عن التوزع Z فى كون أن تباينه يعتمد على حجم العينة n ، وهو أكبر من الواحد دوما . ويصبح التوزيعان متماثلين فقط فى الحالة التى يسعى فيها $n \to \infty$.

ويوضح لنا الشكل (٥,١١) درجة الصلة بين التوزيع t من أجل $\nu = 3,6$ ، والتوزيع الطبيعى المعيارى .

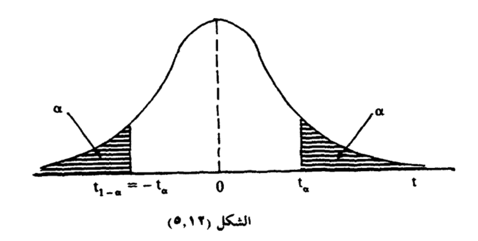


الشكل (٥,١١)

إن احتمال أن تقدم العينة العشوائية قيمة $t = (\bar{X} - \mu) / (S/\sqrt{n})$ واقعة بين قميتين مختلفتين يساوى إلى المساحة الواقعة تحت منحنى التوزيع t بين الاحداثيين السينين لهاتين القيمتين . ويوضح الجدول V القيم t والتى تحدد فوقها (أى على يمينها) مساحة معينة α ، وذلك من أجل بعض القيم α .

$$\alpha = 0.1, 0.05, 0.025, 0.01, 0.005$$

والجدول V مصمم بصورة تختلف عن جدول التوزيع الطبيعى IV . فالسطر الأول يوضح المساحات المختلفة α أما العمود الأيسر فيوضح درجات الحرية . ومن المألوف أن نرمز بـ T_α للقيمة t التى تحدد فوقها (أى على يمينها) مساحة قدرها α . لذلك فإن قيمة t بـ 15 درجة من الحرية والتى تحدد فوقها مساحة قدرها 0.005 هى القيمة $t = 2.947$. وبما أن التوزيع t متناظر حول الوسط المساوى للصفر ، لذلك فإننا نجد أن $t_{1-\alpha} = -t_\alpha$ ، وهذه هى قيمة t التى تحدد لنا مساحة على يمينها قدرها $1 - \alpha$ ، ولذلك فإن قيمة t التى تحدد على يسارها مساحة قدرها α تساوى ناقص القيمة t التى تحدد على يمينها نفس المساحة . لاحظ الشكل (٥,١٢) .



الشكل (٥,١٢)

نلاحظ فى التوزيع t أنه من أجل 15 = أن :

$$t_{0.995} = -t_{0.005} = -2.947$$

وهذا يعنى أن القيمة t من أجل عينة عشوائية حجمها $n = 16$ ومختارة من مجتمع طبيعى ستقع بين القيمتين 2.947, 2.947 − باحتمال قدره 99% .

وبالضبط فإن 99% من التوزيع t بـ (n − 1) درجة من الحرية سيقع بين النقطتين $-t_{0.005}, t_{0.005}$ - لذلك فإن القيمة t الواقعة قبل $-t_{0.005}$ أو أعلى النقطة $t_{0.005}$ ستجعلنا نميل إلى الاعتقاد فى أن حادثاً نادراً سيقع ، أو أن افتراضا حول قيمة μ قد ارتكبنا فيه خطأ .

إن اهتمامنا بـ μ سيحدد لنا طول المجال الذى نقبل فيه قيم t . وبعبارة أخرى إذا كنا نجهل قيمة الوسط الحقيقى والمختلفة بشكل تافه عما ندعيه أن تكون ، فعلينا اختيار مجال بعرض من $-t_{0.01}$ إلى $t_{0.01}$ ، وعلى t أن تقع عندئذ فى هذا المجال . هذا وأن قيمة t الواقعة بالقرب من أحد أطراف المجال ستقودنا إلى الاعتقاد بأن فرضيتنا حول القيمة μ صحيحة . كذلك من المرجح أن تكون بعض القيم القريبة من μ ممثلة للقيمة الحقيقية ، وإذا كانت μ معلومة بدرجة عالية من الدقة ، فيجب علينا عندئذ أن نستخدم أى مجال قصير مثل $(-t_{0.05}, t_{0.05})$ ، وفى هذه الحالة فإن قيمة t الواقعة خارج المجال ستتركنا نعتقد أن القيمة المفروضة لـ μ فيها خطأ ، فى حين أن هذا صحيح بكل معنى الكلمة .

مثال (٥,١٩)

تدعى شركة معينة لإنتاج اللمبات الكهربائية أن لمباتها ستحترق عند مضى 700 ساعة فى المتوسط على عملها . وللتأكد من هذا الإدعاء يقوم مهندسو هذه الشركة بإشعال 30 لمبة كل شهر ويقتنعوا بصحة إدعائهم إذا كانت القيمة المحسوبة لـ t واقعة فى المجال $(-t_{0.05}, t_{0.05})$. ما هو القرار النهائى الممكن استنتاجه من خلال عينة مسحوبة من إنتاج المصنع إذا كان لها وسط $\overline{X} = 715$ ساعة ، وانحراف معيارى $S = 38$ ساعة ؟ افرض أن توزيع احتراق لمبة توزيع قريب من التوزيع الطبيعى .

الحل

من الجدول ٧ نجد أن $t_{0.05} = 1.699$ من أجل $29 = n - 1 = \nu$ درجة من الحرية . لذلك فإن مهندسوا الشركة يقتنعون بصحة فرضيتهم إذا قدمت عينة عشوائية حجمها $n = 30$ من إنتاج الشركة قيمة لـ t محصورة فى المجال $(-1.699, 1.699)$. فإذا كان $\mu = 700$ فعندئذ نجد أن :

$$t = \frac{715 - 700}{38 / \sqrt{30}} = 2.162$$

وهذه القيمة ستكون أكبر من 1.699 إن احتمال الحصول على قيمة لـ t أكبر أو يساوى من 2.162 يساوى تقريبا 0.02 وإذا كانت $\mu > 700$ فعندئذ ستكون قيمة t المحسوبة من خلال عينة معقولة بشكل أكبر . لذلك فإنه من المرجح أن تدعى الشركة أن إنتاجها من اللمبات سيكون أفضل مما اعتقدت .

(٥,٨) التوزيع F

إن أحد التوزيعات فى التطبيقات الإحصائية هو التوزيع F . ويعرف التوزيع F على أنه يمثل نسبة متغيرين عشوائيين (من نوع كاى - مربع) مستقلان ، وكل متغير بدرجة حرية معينة . لذلك فإننا نكتب :

$$F = \frac{U / \nu_1}{V / \nu_2}$$

حيث يمثل U, V متغيرين عشوائيين مستقلين لهما توزيع كاى - مربع بعدد من درجات الحرية مساو لـ ν_1, ν_2 على الترتيب . سنبحث فيما يلى عن الكثافة الاحتمالية للمتغير F .

نظرية (٥,١٨)

بفرض أن U و V متغيران عشوائيان مستقلان لهما توزيع كاى - مربع بعدد من درجات الحرية ν_1 , ν_2 على الترتيب . عندئذ يكون التوزيع الاحتمالى للمتغير :

$$F = \frac{U / \nu_1}{V / \nu_1}$$

من الشكل :

$$h(f) = \begin{cases} \dfrac{[\Gamma(\nu_1 + \nu_2)/2]\,(\nu_1/\nu_2)^{\nu_1/2}}{\Gamma(\frac{\nu_1}{2}) \cdot \Gamma(\frac{\nu_2}{2})} \cdot \dfrac{f^{\nu_1/2 - 1}}{(1 + \frac{\nu_1 f}{\nu_2})^{\frac{\nu_1 + \nu_2}{2}}} & 0<f<\infty \\ 0 & \text{فيما عدا ذلك :} \end{cases}$$

تدعى الدالة h(f) بدالة كثافة التوزيع F بدرجتى الحرية $\nu_1 - \nu_2$

البرهان

من المعلوم أن التوزيع الاحتمالي المشترك للمتغيرين المستقلين u و v يعطى بالعلاقة :

$$(u, v) = f_1(u) \cdot f_2(v)$$

حيث يمثل كلا من $f_1(u)$,$f_2(v)$ كثافة المتغيرين v , u على الترتيب ، لذلك فإن :

$$\phi(u, \nu) = \frac{1}{2^{\frac{\nu_1}{2}} \Gamma(\frac{\nu_1}{2})} \cdot u^{\frac{\nu_1}{2}-1} \cdot e^{-\frac{u}{2}} \frac{1}{2^{\frac{\nu_2}{2}} \Gamma(\frac{\nu_2}{2})} v^{\frac{\nu_2}{2}-1} \cdot e^{-\frac{v}{2}}$$

$$= \begin{cases} \dfrac{1}{2^{\frac{\nu_1+\nu_2}{2}} \Gamma(\frac{\nu_1}{2}) \cdot \Gamma(\frac{\nu_1}{2})} u^{\frac{\nu_1}{2}-1} \cdot V^{\frac{\nu_2}{2}-1} e^{-\frac{1}{2}(u+v)} & : 0 < u < \infty \\ & \quad 0 < n < \infty \\ 0 & \text{فيما عدا ذلك :} \end{cases}$$

لنعرف الآن متغيرا جديدا $w = v$. نلاحظ أن الحل العكسي للعلاقتين $f = \frac{U / \nu_1}{V / \nu_2}$ ، $V = w$ هما : $u = \frac{\nu_1}{\nu_2} f.w$ ، $v = w$ ومنها نحصل على معين جاكوبي :

$$J = \begin{vmatrix} \frac{\nu 1}{\nu 2} w & \frac{\nu 1}{\nu 2} f \\ 0 & 1 \end{vmatrix} = \frac{\nu_1}{\nu_2} w$$

نلاحظ أن التطبيق بين مجموعة النقاط $\{(u, v) | 0<u<\infty , 0<v<\infty\}$ ومجموعة النقاط : $\{(f,w) | 0<f<w , 0<w<\infty\}$ هو تطبيق واحد لواحد . وباستخدام النظرية (٥,٤) نجد أن التوزيع الاحتمالي المشترك للمتغيرين W, F هو من الشكل :

$$g(f, w) = \begin{cases} \dfrac{w^{\nu_2/2-1}}{2^{(\nu_1+\nu_2)/2} \Gamma \frac{\nu_1}{2} \Gamma \frac{\nu_2}{2}} \left(\dfrac{\nu_1 f w}{\nu_2}\right)^{\nu_1/2-1} . e^{-\frac{w}{2}\left[\frac{\nu_1}{\nu_2} f + 1\right]} \left(\dfrac{\nu_1}{\nu_2} w\right)^2 & \begin{matrix} 0 < f < \infty \\ 0 < w < \infty \end{matrix} \\ & \text{فيما عدا ذلك :} \end{cases}$$

أما توزيع المتغير F فنحسبه بواسطة التوزيع الهامشى :

$$h(f) = \int_0^\infty g(f, w)\, dw$$

$$= \frac{1}{2^{\frac{\nu_1+\nu_2}{2}}\, \Gamma(\frac{\nu_1}{2}) \cdot \Gamma(\frac{\nu_2}{2})} \int_0^\infty w^{\left(\frac{\nu_1+\nu_2}{2} - 1\right)} \cdot e^{-\frac{w}{2}\left[\frac{\nu_1}{\nu_2} f + 1\right]}\, dw$$

وباجراء تبديل فى المتحول بالعلاقة : $Z = \frac{W}{2}\left[\frac{\nu_1}{\nu_2} f + 1\right]$ نجد أن :

$$dW = \frac{2}{(\frac{\nu_1}{\nu_2} f + 1)}\, dz$$

ومنه :

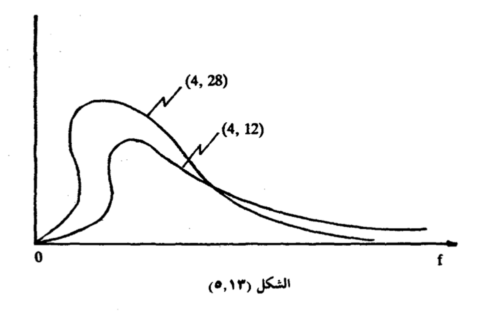
$$h(f) = \frac{(\frac{\nu_1}{\nu_1})^{\frac{\nu_1}{2}} f^{\frac{\nu_1}{2}-1}}{2^{\frac{\nu_1+\nu_2}{2}} \cdot \Gamma(\frac{\nu_1}{2}) \cdot \Gamma(\frac{\nu_2}{2})}$$

$$\int_0^\infty \left(\frac{2Z}{\frac{\nu_1 f}{\nu_2 + 1}}\right)^{\frac{\nu_1+\nu_2}{2} - 1} \cdot e^{-z} \frac{2}{(\frac{\nu_1 f}{\nu_2 + 1} + 1)}\, dz$$

$$= \frac{(\frac{\nu_1}{\nu_2})^{\frac{\nu_1}{2}} f^{\frac{\nu_1}{2}-1}}{\Gamma(\frac{\nu_1}{2}) \cdot \Gamma(\frac{\nu_2}{2}) \cdot (1 + \frac{\nu_1}{\nu_2} f)^{\frac{\nu_1+\nu_2}{2}}} \int_0^\infty z^{\frac{\nu_1+\nu_2}{2} - 1}\, e^{-z}\, dz$$

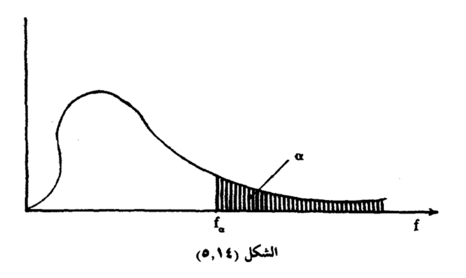
$$= \begin{cases} \dfrac{\Gamma(\frac{\nu_1+\nu_2}{2})\ (\frac{\nu_1}{\nu_2})^{\frac{\nu_1}{2}}\ f^{\frac{\nu_1}{2}-1}}{\Gamma(\frac{\nu_1}{2}) \cdot \Gamma(\frac{\nu_2}{2})\ (1 + \frac{\nu_1}{\nu_2} f)^{\frac{\nu_1+\nu_2}{2}}} & : \ 0 < f < \infty \\ 0 & \text{فيما عدا ذلك :} \end{cases}$$

نلاحظ فى عبارة F أن درجة حرية المتغير الأول u قد وردت فى بسط الكسر . كما أن درجة حرية المتغير الثانى V قد وردت فى مقام الكسر . وهذا يعنى أن منحنى توزيع المتغير F لا يتعلق بدرجتى حرية المتغيرين U و V وإنما يتعلق أيضا بالترتيب الذى نبدأ فيه . ويوضح الشكل (٥,١٣) منحنيين للتوزيع F من أجل درجات الحرية , (4, 12) . (4,28)



الشكل (٥,١٣)

لنفرض أن f_α تمثل قيمة خاصة للمتغير العشوائى F ، والتى تحدد لنا على يمينها مساحة قدرها α يوضحها القسم المظلل على الشكل (٥,١٤) الموضح أدناه .



الشكل (٥,١٤)

ويوضح لنا الجدول VII القيم f_α من أجل $\alpha = 0.05$ ، $\alpha = 0.01$ وذلك من أجل عدد من الأزواج $({}_1 , {}_2)$. نلاحظ أن قيمة f بـ (4, 12) درجات من الحرية والتى تحدد لنا على يمينها مساحة قدرها 0.05 هى القيمة $f_{0.05} = 3.26$. وتوضح لنا النظرية التالية طريقة لاستخدام الجدول VII فى حساب قيم $f_{0.95}$ ، $f_{0.99}$.

نظرية (٥,١٩)

بفرض أن $f_\alpha ({}_1 , {}_2)$ تمثل القيمة f_∞ بدرجتى الحرية $({}_1 , {}_2)$ عندئذ :

$$f_{1-\alpha}(\nu_1, \nu_2) = \frac{1}{f_\alpha(\nu_1 , \nu_2)}$$

وهكذا فإن قيمة f بدرجتى الحرية (4,12) والتى تحدد لنا على يمينها مساحة قدرها 0.95 هى :

$$f_{0.95}(4, 12) = \frac{1}{f_{0.05}(12, 4)} = \frac{1}{5.91} = 0.169$$

لنفرض أننا سحبنا عينتين عشوائيتين حجمهما على الترتيب n_1, n_2 من مجتمعين طبيعيين بالتباينين σ_1^2, σ_2^2 على التتالى . من النظرية (٥,١٦) نعلم أن لكل من المتغيرين :

$$\chi_1^2 = \frac{(n_1 - 1)\, S_1^2}{\sigma_1^2}$$

$$\chi_2^2 = \frac{(n_1 - 1)\, S_2^2}{\sigma_2^2}$$

توزيعا من نوع كاى – مربع بـ $\nu_1 = n_1 - 1$ ، $\nu_2 = n_2 - 1$ على الترتيب درجات من الحرية . علاوة على ذلك ، بما أن العينات عشوائية ، لذلك فإنها ستكون مستقلة ، وباستخدام النظرية (٥,١٨) بعد فرض أن :

$$\chi_1^2 = U, \chi_2^2 = V$$

نجد النظرية التالية :

نظرية (٥,٢٠)

بفرض أن S_1^2 , S_2^2 يمثلان تبايني عينتين مستقلتين حجماهما على الترتيب n_1 , n_2 مأخوذتين من مجتمعين طبيعيين بالتباينين σ_1^2 , σ_2^2 فعندئذ يكون للمتغير :

$$F = \frac{S_1^2 / \sigma_1^2}{S_2^2 / \sigma_2^2} = \frac{\sigma_2^2 S_1^2}{\sigma_1^2 S_2^2}$$

توزيعا من نوع F بعدد من درجات الحرية مساوى لـ :

$$(v_1 , v_2) = (n_1 - 1 , n_2 - 1)$$

فى الفصل السادس سنستخدم النظرية (٥,٢٠) لإيجاد ثقة لتبايني مجتمعين طبيعيين ، يستخدم التوزيع F على نطاق واسع فى طرق الإحصاء . ونلاحظ أن $t^2 = \frac{Z^2}{V/\nu}$ هو نسبة $\chi^2(1)$ إلى $\frac{\chi^2(\nu)}{\nu}$ أى أن توزيع المتغير $[t(\nu)]^2$ هو بالتعريف $F(1, \nu)$ ، وهذا يعنى أن التوزيع $F(1,\nu)$ يكافىء مربع التوزيع $t(\nu)$.

تمارين محلولة

تمرين (١)

بفرض X متغيراً عشوائيا حدانيا . توزيعه الاحتمالى من الشكل :

$$f(x) = \begin{cases} \binom{3}{x} \cdot \left(\frac{2}{5}\right)^{x}\left(\frac{2}{5}\right)^{3-x} & : x = 0, 1, 2, 3 \\ 0 & : \text{فيما عدا ذلك} \end{cases}$$

ما هو التوزيع الاحتمالى للمتغير الجديد $Y = X^2$ ؟

الحل

نلاحظ أن المتغير المفروض X منقطع وقيمه جميعها موجبة ، لذلك فالتقابل $Y = X^2$ هو تقابل واحد لواحد بين مجموعة قيم x وقيم y ، وحسب النظرية (١،٥) نستطيع أن نكتب بعد حل المعادلة $y = x^2$ بالنسبة لـ x .

$$g(y) = f[(\sqrt{y})]$$

$$= \begin{cases} \binom{3}{\sqrt{y}} \left(\frac{2}{5}\right)^{\sqrt{y}} \left(\frac{3}{5}\right)^{3-\sqrt{y}} & : y = 0, 1, 4, 9 \\ 0 & : \text{فيما عدا ذلك} \end{cases}$$

تمرين (٢)

بفرض أن التوزيع الاحتمالى المشترك X_1, X_2 هو من الشكل :

$$f(x_1, x_2) = \binom{2}{x_1, x_2, 2-x_1-x_2} \cdot \left(\frac{1}{4}\right)^{x_1} \cdot \left(\frac{1}{3}\right)^{x_2} \cdot \left(\frac{5}{12}\right)^{2-x_1-x_2}$$

وذلك من أجل $x_1 = 0,1,2$ ، $x_2 = 0,1,2$ ، $x_1 + x_2 \leq 2$ وصفر فيما عدا ذلك :

ما هو التوزيع المشترك للمتغيرين $Y_1 = X_1 + X_2$ و $Y_2 = X_1 - X_2$ ؟

الحل

من الواضح أن التطبيقين Y_1 ، Y_2 هما تطبيقان واحد لواحد بين مجموعة النقاط (y_1, y_2) و مجموعة النقاط (Y_1, Y_2) وحسب النظرية (٥،٢٠) نكتب[بعد حل المعادلتين $y_2 = x_1 - x_2$, $y_1 = x_1 + x_2$ حلا عكسيا بالنسبة لكل من x_2 و x_1 نجد أن :

$$\left[x_1 = \frac{y_1 + y_2}{2}\ ,\ x_2 = \frac{y_1 - y_2}{2}\right.$$

$$g(y_1, y_2) = f\left[\frac{y_1 + y_2}{2}, \frac{y_1 - y_2}{2}\right]$$

$$g(y_1, y_2) = \begin{cases} \binom{2}{\frac{y_1 + y_2}{2},\ \frac{y_1 - y_2}{2},\ 2 - y_1} \left(\frac{1}{4}\right)^{\frac{y_1 + y_2}{2}} \left(\frac{1}{3}\right)^{\frac{y_1 - y_2}{2}} \cdot \left(\frac{5}{11}\right)^{2 - y_1} & y_1 = 0, 1, 2 \\ & y_2 = -1, 0, 1 \\ 0 & \text{فيما عدا ذلك :} \end{cases}$$

وهكذا نجد أن :

$$g(y_1, y_2) = \begin{cases} \binom{2}{\frac{y_1 + y_2}{2},\ \frac{y_1 - y_2}{2},\ 2 - y_1} \left(\frac{121}{300}\right)^{\frac{y_1}{2}} \cdot \left(\frac{3}{4}\right)^{\frac{y_2}{2}} \left(\frac{5}{11}\right)^{2} : & : y_1 = 0, 1, 2,\ y_2 = -1\ 0, 1 \\ & y_2 \leq y_1,\ y_1 + y_2 \text{ are even} \\ 0 & \text{فيما عدا ذلك :} \end{cases}$$

تمرين (٣)

بفرض أن X_1, X_2 يمثلان متغيرين عشوائيين منقطعين بالتوزيع الاحتمالي المشترك :

$$f(x_1, x_2) = \begin{cases} \frac{x_1 x_2}{18} & : x_1 = 1, 2 \\ & \quad x_2 = 1, 2, 3 \\ 0 & \text{فيما عدا ذلك :} \end{cases}$$

أوجد التوزيع الاحتمالى للمتغير $Y = X_1 . X_2$

الحل

لنفتش أولاً عن التوزيع الاحتمالى المشترك للمتغيرين Y , X_2 . فحسب النظرية (٥,٢) وبملاحظة أن كلا من التقابلين Y , X_2 هو واحد لواحد فإننا نجد أن :

$$g(y, x_2) = f\left[\frac{y}{x_2} . x_2\right]$$

$$= \begin{cases} \dfrac{\frac{y}{x_2} . x_2}{18} & : y = 1,2,3,4,6 \\ & \quad x_2 = 1,2,3, y / x_2 = 1 , 2 \\ 0 & : \text{فيما عدا ذلك} \end{cases}$$

وبالجمع على قيم $x_2 = 1, 2, 3$ نجد التوزيع الهامشى للمتغير y ومنه :

$$g(y) = \begin{cases} \dfrac{y}{18} & : y = 1,3,4,6 \\ \dfrac{2y}{18} & : y = 2 \\ 0 & : \text{فيما عدا ذلك} \end{cases}$$

تمرين (٤)

إذا علمت أن للمتغير العشوائى المستمر X توزيعا طبيعيا بالوسط $\mu = 0$ والتباين σ^2 فما هى دالة الكثافة للمتغير $Y = aX^2$ وحيث أن $a > 0$ ؟

الحل

لقد برهنا فى التمرين (٥,٥) أنه إذا كان للمتغير x توزيعا طبيعيا بالمتغيرين μ , σ كان للمتغير $\left(\frac{X - \mu}{\sigma}\right)^2$ توزيعا من نوع كاى – مربع بدرجة واحدة من الحرية وبالعودة إلى التمرين المذكور وبفرض أن $\mu = 0$ نجد أن للمتغير $Z = \frac{X^2}{\sigma^2}$ توزيعا يشبه توزع كاى – مربع بدرجة واحدة من الحرية أى أن للمقدار X^2 توزيعا من الشكل :

$$f_{\chi^2}(z) = \begin{cases} \dfrac{1}{\sqrt{2\pi z}\ .\ \sigma}\, e^{-\frac{1}{2}\frac{z}{\sigma^2}} & : z>0 \\ 0 & : \text{فيما عدا ذلك} \end{cases}$$

ومن المعلوم حسب النظرية (٥,٣) أن :

$$f_Y(y) = f_{\chi^2}(y / a)\ .\ \left[\frac{d\,(x^2)}{dy} \right]$$

محسوبا عند علاقة التحويل $x = \frac{y}{a}$

لذلك فإن :

$$f_Y(y) = \begin{cases} \dfrac{1}{\sqrt{2\,\pi\, a\,.\, y}\,\sigma}\, e^{-\frac{1}{2}\frac{y}{a\,.\,\sigma^2}} & : y > 0 \\ 0 & : \text{فيما عدا ذلك} \end{cases}$$

تمرين (٥)

بفرض أن دالة كثافة الثنائية $(x\,,\,y)$ هي من الشكل :

$$f\,(x\,,\,y) = \quad (x\,,\,y) \quad .\,|\,J\,|$$

$$= \frac{1}{\sigma_x\,\sigma_y\,2\,\pi}\, e^{\ \frac{1}{2}\left[\frac{X^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right]}$$

فما هي دالة كثافة المتغير $z = x - y$ ؟

الحل

لنبحث أولا عن دالة الكثافة للثنائية (z, y)

فحسب النظرية (٥,٤) يمكن أن نكتب :

$$f(z, y) = f(z + y, y)\ .\,|\,J\,|$$

$$= \frac{1}{\sigma_x\,\sigma_y\,2\,\pi}\, e^{-\frac{1}{2}\left[\frac{(z + y)^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right]}$$

ذلك لأن :

$$|J| = \begin{vmatrix} \frac{\partial x}{\partial z} & \frac{\partial x}{\partial y} \\ \frac{\partial y}{\partial z} & \frac{\partial y}{\partial y} \end{vmatrix} = \begin{vmatrix} 1 & 0 \\ \frac{\partial y}{\partial z} & 1 \end{vmatrix} = 1$$

وبعد مكاملة الدالة الأخيرة بالنسبة لـ y نجد دالة الكثافة للمتغير Z هى .

$$f(z) = \frac{1}{\sigma_x \sigma_y 2\pi} \int_{y=-\infty}^{+\infty} e^{-\frac{1}{2}\left[\frac{(z+y)^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right]} dy$$

$$= \frac{e^{-\frac{1}{2}\frac{z^2}{\sigma_x^2}}}{\sigma_x \sigma_y 2\pi} \int_{y=-\infty}^{+\infty} e^{-\frac{1}{2}\left[y^2\left(\frac{\sigma_x^2 + \sigma_y^2}{\sigma_x^2 . \sigma_y^2}\right) + 2yz\,\sigma_y^2\right] / \sigma_x^2 . \sigma_y^2}$$

لنتم لمربع كامل فنجد أن :

$$= \frac{e^{-\frac{z^2}{2}\left(\frac{1}{\sigma_x^2 = \sigma_y^2}\right)}}{\sigma_2 \sqrt{2\pi}}$$

حيث إن : $\sigma_z = \sqrt{\sigma_1^2 + \sigma_2^2}$

تمرين (٦)

بفرض أن التوزيع المشترك للمتغيرين العشوائيين المستمرين X, Y هو من الشكل :

$$f(x, y) = \frac{1}{2\pi . \sigma^2} e^{-\frac{x^2 + y^2}{2\sigma^2}}$$

وبفرض أن $X = R . \cos\theta$, $Y = R \sin\theta$

فما هو التوزيع المشترك للمتغيرين R, θ ؟

الحل

حسب النظرية (٥,٤) نكتب أن دالة الكثافة للمتغيرين R, هى من الشكل :

$$f(r,\theta) = f(x,y) . \begin{vmatrix} \frac{\partial x}{\partial r} & \frac{\partial y}{\partial r} \\ \frac{\partial x}{\partial \theta} & \frac{\partial y}{\partial \theta} \end{vmatrix}$$

محسوبة عند $y = r\sin\theta \;,\; x = r\cos\theta$

ومنه :

$$f(\nu, \theta) = \frac{r}{2\pi\sigma^2} e^{-\frac{1}{2\sigma^2} r^2} : \begin{array}{l} 0 < r < \infty \\ 0 \leq \theta \leq 2\pi \end{array}$$

تمرين (٧)

برهن أن الدالة المولدة للعزوم لمتغير عشوائى هندسى هى من الشكل :

$$M_x(t) = \frac{p\,e^t}{1 - qe^t}$$

البرهان :

من المعلوم أن التوزيع الاحتمالى للمتغير الهندسى يعطى بالجدول :

X	1	2	3	...	x	...
f(x)					$p \cdot q^{x-1}$	

وحسب التعريف (٥,١) نجد أن :

$$M_x(t) = E[e^{tx}] = \Sigma\, e^{tx} . f(x)$$

$$= \sum_{x=1}^{\infty} e^{tx} . p . q^{x-1}$$

$$= \frac{p}{q} \sum_{x=1}^{\infty} (e^t . q)^x$$

$$= \frac{P}{q} \cdot [\sum_{x=0}^{\infty} (e^{t} \cdot q)^{x}] \cdot q \cdot e^{t}$$

ومن الواضح أن $|e^{t} . q| \leq 1$

$$\sum_{x=0}^{\infty} (qe^{t})^{x} = \frac{1}{1 - qe^{t}}$$

لذلك فإن :

$$M_{x}(t) = \frac{p\, e^{t}}{1 - qe^{t}}$$

تمرين (٨)

أوجد الدالة المولدة للعزوم لمتغير عشوائي بواسونى X بالوسط μ .

الحل

نعلم أن التوزيع الاحتمالى لمتغير بواسون بالوسط μ هو من الشكل :

X	0	1	2	...	x	...
f(x)					$e^{-\mu} \frac{\mu^{x}}{X!}$	

وبالعودة إلى التعريف (٥,١) فإننا نجد أن :

$$M_{x}(t) = \sum_{x=0}^{\infty} e^{tx} \cdot \frac{e^{-\mu} \cdot \mu^{x}}{X!}$$

$$= e^{-\mu} \cdot \sum_{x=0}^{\infty} \frac{(\mu \cdot e^{t})^{x}}{X!}$$

ومن المعلوم أن :

$$\sum_{x=0}^{\infty} \frac{\alpha^{x}}{x!} = e^{\alpha}$$

إذن:

$$M_x(t) = e^{-\mu} \cdot e^{\mu e^t}$$
$$= e^{(e^t - 1)\mu}$$

وهى الدالة المولدة للعزوم لمتغير بواسونى بالوسط μ .

تمرين (٩)

مستخدما الدالة المولدة للعزوم فى التمرين (٥,٨) بيِّن أن توقع وتباين متغير كاى – مربع بـ ν درجة حرية هما على الترتيب .

البرهان

من التمرين (٥,٨) نجد أن :

$$M_x(t) = \frac{1}{(1 - 2t)^{\nu/2}}$$

وحسب النظرية (٥,٦) نعلم أن :

$$\mu = \left.\frac{dM_x(t)}{dt}\right|_{t=0} , \quad \mu_2 = \left.\frac{d^2M_x(t)}{dt^2}\right|_{t=0}$$

وباشتقاق الدالة السابقة مرتين بالنسبة لـ t فإننا نجد أن :

$$\frac{dM_x(t)}{dt} = -\frac{\nu}{2}(-2)(1 - 2t)^{-\frac{\nu}{2}-1} = \frac{\nu}{(1 - 2t)^{\frac{\nu}{2}+1}}$$

$$\frac{d^2M_x(t)}{dt^2} = -(\frac{\nu}{2} + 1)(v)(-2)(1 - 2t)^{-\frac{\nu}{2}-2}$$

ويوضع $t = 0$ فى العلاقتين السابقتين فإننا نجد :

$$\mu = \nu \quad , \quad \mu_2 = 2\nu\left(\frac{\nu}{2} + 1\right)$$

ومنه :

$$\sigma^2 = \mu'^2 - \mu^2 = 2\nu\left(\frac{\nu}{2} + 1\right) - \nu^2 = 2\nu$$

تمرين (١٠)

استخدم الجدول VI فى إيجاد :

(١) $\chi^2_{0.01}$ بـ $\nu = 18$

(٢) $\chi^2_{0.975}$ بـ $\nu = 29$

(٣) χ^2_{α} بحيث تتحقق المعادلة $P[X^2 < \chi^2_{\alpha}] = 0.99$ وذلك من أجل $\nu = 4$:

الحل

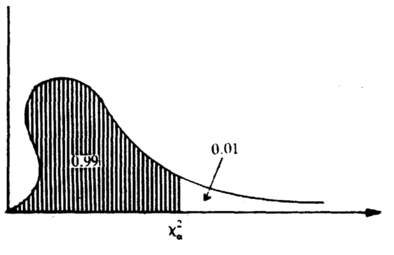
من الجدول VI نجد أن :

$$\chi^2_{0.01} = 34.805 \quad : \quad \nu = 18$$
$$\chi^2_{0.975} = 16.047 \quad : \quad \nu = 29$$

ولإيجاد قيمة χ^2_{α} التى تحدد على يسارها مساحة قدرها 0.99 علينا أن نبحث فى الجدول المذكور عن χ^2_{α} التى تحدد على يمينها مساحة قدرها 0.01 من أجل $\nu = 4$. لذلك فإننا نجد أن قيمة χ^2_{α} المطلوبة والمحققة للعلاقة :

$$P[X < \chi^2_{\alpha}] = 0.99$$

هى $\chi^2_{0.01}$ ومنه $\chi^2_{\alpha} = 13.277$ انظر الشكل المرفق :



تمرين (١١)

احسب احتمال أن يكون تباين عينة S^2 حجمها $n = 25$ مسحوبة من مجتمع طبيعى بالتباين $\sigma^2 = 6$

١) أكبر من 9.1

٢) بين القيمتين (3.462, 10.745)

الحل

أولا ــ نعلم أن :

$$P[S^2 > 9.1] = P\left[\frac{(n-1)S^2}{\sigma^2} > \frac{24}{6}(9.1)\right]$$

$$= P[\chi^2_{(24)} > 36.4]$$

لنبحث الآن عن المساحة التى يحددها العدد 36.4 على يمينه فى التوزيع كاى – مربع بـ $\nu = 24$ درجة من الحرية . من الجدول VI نجد أن هذا الاحتمال يمثل المساحة 0.05 .

ثانيا ــ نعلم أن :

$$P[3.462 < S^2 < 10.745] = P\left[13.848 < \frac{(n-1)S^2}{\sigma^2} < 42.98\right]$$

$$= P[13.848 < \chi^2_{(24)} < 42.98]$$

$$= \chi^2_{13.848} - \chi^2_{42.98}$$

$$= 0.95 - 0.01$$

$$= 0.94$$

تمرين (١٢)

باستخدام الجدول V أوجد :

(١) $t_{0.025}$ من أجل $\nu = 17$

(٢) $t_{0.09}$ من أجل $\nu = 10$

(٣) t_α بحيث تتحقق العلاقة $P[-t_\alpha < T < t_\alpha] = 0.90$
وذلك من أجل $\nu = 23$

الحل

من الجدول V نجد أن :

$$t_{0.025} = 2.11 \quad , \quad \nu = 17$$

ومن المعلوم أن $t_\alpha = -t_{1-\alpha}$ لذلك فإنه من أجل 10 =

$$t_{0.99} = -t_{0.01} = -2.764$$

ثالثا ــ نلاحظ أن t_α المطلوب حسابها يجب أن تحدد علي يمينها مساحة قدرها 0.05 ، وذلك لأن توزيع t متناظر حول الوسط الذى يساوى صفراً . وهكذا نجد أن :

$$t_{0.05} = 1.714$$

وذلك من أجل $\nu = 23$.

تمارين عامة

(١) ليكن X متغيراً عشوائياً منقطعاً بالتوزيع الاحتمالى :

$$f(x) = \begin{cases} \frac{1}{3} & : x = 1, 2, 3 \\ 0 & : \text{فيما عدا ذلك} \end{cases}$$

أوجد التوزيع الاحتمالى للمتغير $Y = 3X - 3$

(٢) بفرض أن X يمثل متغيراً عشوائياً حدانياً بالتوزيع الاحتمالى :

$$f(x) = \begin{cases} \binom{3}{x} \left(\frac{2}{5}\right)^x \left(\frac{3}{5}\right)^{3-x} & : x = 0, 1, 2, 3 \\ 0 & : \text{فيما عدا ذلك} \end{cases}$$

أوجد التوزيع الاحتمالى للمتغير $Y = X^3$

(٣) ليكن X_1 , X_2 متغيرين عشوائيين منقطعين بالتوزيع المشترك :

$$f(x_1, x_2) = \begin{cases} \binom{2}{x_1, x_2, 2 - x_1 - x_2} \left(\frac{1}{4}\right)^{x_1} \left(\frac{1}{3}\right)^{x_2} \left(\frac{5}{12}\right)^{2-x_1-x_2} & : x_1 = 0, 1, 2 \\ & : x_2 = 0, 1, 2 \\ 0 & : \text{فيما عدا ذلك} \end{cases}$$

ما هو التوزيع الاحتمالى المشترك للمتغيرين :

$$Y_1 = X_1 + X_2 \quad , \quad Y_2 = X_1^2 + X_1 . X_2$$

(٤) بفرض أن للمتغير العشوائى المستمر X كثافة احتمالية من الشكل :

$$f(x) = \begin{cases} 1 & : 0 < x < 1 \\ 0 & : \text{فيما عدا ذلك} \end{cases}$$

برهن أن للمتغير $Y = -2 \ln X$ توزيعاً من نوع كاى – مربع بدرجتى حرية .

(٥) يتغير التيار I أمبير المتدفق فى المقاومة R أوم وفقا للتوزيع الاحتمالى :

$$f(i) = \begin{cases} 6i(1-i) & : 0 < i < 1 \\ 0 & : \text{فيما عدا ذلك} \end{cases}$$

فإذا علمت أن المقاومة تتغير بصورة مستقلة عن التيار وفقا للتوزيع الاحتمالى :

$$g(r) = \begin{cases} 2r & : 0 < r < 1 \\ 0 & : \text{فيما عدا ذلك} \end{cases}$$

فما هو التوزيع الاحتمالى للقوة $W = I^2.R$ واط ؟

(٦) بفرض أن التوزيع المنتظم للمتغير المنقطع X هو من الشكل :

$$f(x) = \begin{cases} \frac{1}{n} & : x = 1, 2, 3, \dots, n \\ 0 & : \text{فيما عدا ذلك} \end{cases}$$

أ ــ برهن أن الدالة المولدة للعزوم للمتغير X هى من الشكل :

$$M_x(t) = \frac{e^t(1 - e^{nt})}{n(1 - e^t)}$$

ب ــ باستخدام الدالة $M_x(t)$ السابقة . احسب توقع وتباين المتغير X السابق .

(٧) استخدم الدالة المولدة للعزوم فى التمرين (٦) فى حساب توقع وتباين المتغير الهندسى .

(٨) إذا علمت أن الدالة المولدة للعزوم لمتغير بواسونى معين هى من الشكل :

$$M_x(t) = e^{4(e^t - 1)}$$

فأوجد :

$$P(\mu - 2\sigma < X < \mu + 2\sigma)$$

(٩) سجلنا عدد أيام الغياب لثمانية عشر طالبا فى كلية الهندسة خلال العام الفائت فوجدنا أنها تشكل عينة عشوائية من الشكل :

1,3,4,0,4,2,3,1,2,3,0,4,1,1,1,5,1,0

احسب وسط العينة .

(١٠) احسب تباين العينة 3,5,8,7,5,7

(١١) أوجد تباين العينة 6,10,16,14,10,14 بدون حساب .

(١٢) سحبنا عينة حجمها $N = 16$ من مجتمع طبيعى وسطه $\mu = 50$ وتباينه $\sigma^2 = 25$.
أوجد احتمال أن يقع وسط العينة $\bar{X}$ ضمن المجال :

$$(\sigma_{\bar{x}}^2 - 1.9\mu_{\bar{x}} \ , \mu_{\bar{x}} + 0.4\sigma_{\bar{x}})$$

وذلك بفرض أنه يمكن قياس وسط العينى بأى درجة من الدقة .

(١٣) بفرض أن لأطوال ألف طالب توزيعا قريبا من التوزيع الطبيعى بوسط قدره 174.5 سنتيمتر وانحراف معيارى قدره 6.9 سنتيمتر . لنسحب مئتى عينة مؤلفة من خمسة وعشرين طالبا وبشكل عشوائى من هذا المجتمع ولنسجل وسطاء العينات بالنسبة لأقرب عشر من السنتيمتر ما هو :

أ ــ توقع الوسط والانحراف المعيارى لوسط توزيع المعاينة .

ب ــ عين عدد وسطاء العينات الواقعة فى المجال [172.5,175.8] .

جـ ــ عين عدد وسطاء العينات الواقعة أسفل العدد 172 سنتيمتر .

(١٤) تجربة حدانية مؤلفة من $n = 25$ اختباراً فيها $p = 0.4$. احسب $P(8 \leq X \leq 11)$ وذلك :

أ ــ باستخدام جدول التوزيع الحدانى (الجدول II) .

ب ــ باستخدام طريقة التقريب الطبيعى للتوزيع الحدانى .

(١٥) وجد بالتجربة أن الفترة اللازمة لإنجاز اختبار للذكاء مخصص لطلبة كلية الهندسة يتوزع وفقا للتوزيع الطبيعي بوسط $\mu = 65$ دقيقة وانحراف معياري قدره $\sigma = 10$ دقائق . فكم يجب أن يكون زمن الاختبار إذا أردنا إتاحة وقت كاف لـ 80% من الطلاب لإتمام الاختبار ؟

(١٦) يتوزع عمر نوع من الثلاجات الكهربائية وفقا للتوزيع الطبيعي بوسط $\mu = 3.1$ سنة وانحراف معياري $\sigma = 1.2$ سنة . فإذا كانت الثلاجات مكفولة لمدة عام ، فما هي نسبة الثلاجات المباعة والتي ستضطر الشركة المنتجة لاستبدالها بثلاجات جديدة ؟

(١٧) بنشر الدالة الأسية e^{tx} حسب سلسلة ماكلوران ثم مكاملة كل حد من حدودها ، برهن أن :

$$M_x(t) = \int_{-\infty}^{+\infty} e^{tx} \cdot f(x)dx$$

$$= 1 + \mu t + \mu_2 \frac{t^2}{2\,!} + \dots + \mu'_v \frac{t}{v\,!} + \dots$$

(١٨) عينتان عشوائيتان حجمهما على الترتيب $n_1 = 25, n_2 = 36$ مأخوذتين من مجتمعين مختلفين بالوسطين $\mu_1 = 80, \mu_2 = 75$ وانحرافين معيارين $\sigma_2 = 3$ ، $\sigma_1 = 5$ على التتالي .

احسب احتمال أن يكون وسط العينة $\overline{X}_1$ أكبر من وسط العينة $\overline{X}_2$ بـ 3.4 على الأقل ، وأقل من 5.9 .

(١٩) هل من الممكن الحصول على عينة حجمها $n = 9$ ، وسطها $\overline{X} = 24$ ، وانحرافها المعياري $\sigma_{\overline{x}} = 4.1$ من مجتمع طبيعي تباينه مجهول ووسطه $\mu \geq 20$ ؟

(٢٠) باستخدام الجدول VII احسب :

$f_{0.05}$ من أجل $\nu_1 = 7$ $\nu_2 = 19$

$f_{0.01}$ من أجل $\nu_1 = 24$ $\nu_2 = 19$

$f_{0.95}$ من أجل $\nu_1 = 19$ $\nu_2 = 24$

$f_{0.99}$ من أجل $\nu_1 = 28$ $\nu_2 = 12$

(٢١) أعلنت مديرية الإذاعة والتلفزيون أن 20% من المشاهدين يتابعون برنامجا معينا . وقد وجد من خلال عينة عشوائية مؤلفة من ألف مشاهد ، أن 150 شخصا من بينهم يتابعون البرنامج المذكور فهل تقدم هذه المعلومات دلالة كافية لتأييد ما أعلنته مديرية الإذاعة والتليفزيون ؟

(٢٢) تدعى شركة لإنتاج السجاير أن متوسط القطران فى سجائرها 18.3 مليغرام . اخترنا عينة عشوائية من إنتاج هذه الشركة حجمها $n = 8$ سيجارة ، وحللناها وسجلنا وسط القطران الموجود فيها فوجدنا أنه :

20, 17, 21, 19, 22, 21, 20, 16

فهل توافق إدعاء الشركة السابق أم لا ؟

(٢٣) سحبنا عينتين عشوائيتين مستقلتين حجمهما على الترتيب $n_1 = 25, n_2 = 31$ من مجتمعين طبيعيين بالتباينين $\sigma_2^2 = 15$ ، $\sigma_1^2 = 10$ على الترتيب . احسب :

$$P\left(S_1^2 / S_2^2 < 1.26\right)$$

الفصل السادس

نظرية التقدير

■ مقدمة ■ طرق التقدير الكلاسيكية ■ تقدير الوسط ■ تقدير فرق وسطين ■ تقدير P في المجتمع الحداني ■ تقدير الفرق بين نسبتى مجتمعين حدانيين ■ تقدير التباين ■ تقدير نسبة تباينين ■ تمارين محلولة ■ تمارين عامة .



(٦,١) مقدمة

يواجه الفرد خلال حياته اليومية حالات تتطلب منه القيام بتنبؤات حول المستقبل . فنجد مثلا أن شركة صناعية للمبات الكهربائية تستخدم نتائج تجربة معينة لاستقراء ما إذا كان نوعاً جديداً من اللمبات أكثر جودة من نوع آخر . كما تهتم كل حكومة بالتنبؤ بعدد الطلاب في مختلف المراحل خلال السنوات العشر القادمة . ويرغب مهندسو شركة لإنتاج الأسلحة في معرفة ما إذا كان الفولاذ المستورد من بلد معين أكثر مقاومة لتغيرات الحرارة من نوع آخر . ومن حسن الحظ أنه يمكن أن نتخذ قرارا ونقوم بتنبؤ حول الشيء المدروس في كل حالة من خلال بعض المعلومات المتوفرة لنا ، والتي ندعوها بالملاحظات أو المعلومات الإحصائية . وفي العديد من الحالات تكون المعلومات المتوفرة لدى الدارس لحالة معينة غريزية ، منسجمة في ظاهرها ، ومن نواح عديدة على درجة كبيرة من الوصوح بحيث لا يكون التنبؤ المتخذ بعناية أفضل بكثير من مجرد التخمين الشخصي . ومن جهة أخرى نجد أن التحليل الشخصي للمعلومات من قبل المهندسين مثلا يؤدي غالبا إلى آراء متعارضة حول النتائج المستخلصة لتجربة معينة . وفي الوقت الذي يتفق فيه العديد من الناس بالشعور بقدرتهم الذاتية على القيام باستقراءات جيدة ، إلا أن التجربة تدل على أنه ليس باستطاعة الإنسان أن يجهد فكره بكمية كبيرة من الأرقام وتحليل وموازنة القليل من المعلومات للوصول إلى استقرار جيدة . لهذا يصبح وضع قوانين وأدوات للاستقراء أمراً مرغوباً به وهذا هو هدف الإحصاء .

إن هدف الإحصاء هو القيام باستقراءات حول المجتمع المدروس بدءا من معلومات تقدمها العينة . وإلى المدى الذى تتميز فيه المجتمعات المدروسة بمقاييس وصفية رقمية تدعى بالوسطاء ، لذلك فالاستقراء الاحصائى يهتم بتوفير استقراءات جيدة حول هؤلاء الوسطاء . ومن الأمثلة على هؤلاء الوسطاء المتوسط والتشتت .

والقيام باستقراء حول وسيط يمكن أن ينجز بطريقتين . فيمكن أن نقوم بتقدير الوسط أى التنبؤ بقيمته أو أن نتخذ قراراً يتعلق بقيمة الوسيط . ونلاحظ أنه لابد من وجود مقياس بجودة كل طريقة بحيث يمكن مقارنة هذه الطرق ببعضها والقيام بمفاضلة فيما بينها . هذا بالإضافة إلى أننا نريد التعبير عن جودة استقراء معين فى حالة فيزيائية معينة . فالتنبؤ بأن السعر

العالمى للنفط سيكون ٤٥ دولار للبرميل الواحد فى السنة القادمة ليس كافياً ، ويجب ألا يشكل حافزا للدول المستهلكة للشراء أو عدم الشراء .

فالسؤال الأساسى هو ما إذا كان هذا التقدير صحيحا فى حدود خمسة دولارات زيادة أو نقصانا . هذا ويحوى الاستقراء الإحصائى لحالة معينة عنصرين أساسيين لاغنى عنهما وهذين العنصرين هما ١ ـــ الاستقراء ٢ ـــ مقياس جودة هذا الاستقراء . والسؤال المطروح هو : أية طريقة أفضل فى الاستقراء ، التقدير أم اختبار الفرضيات ؟ أما الجواب فيعود إلى طبيعة المسألة المدروسة وإلى التفضيل الشخصى . ولابد من أن نشير إلى أن للتقدير نوعين . فهناك التقدير النقطى والتقدير المجالى . ففى حالة التقدير النقطى يقوم الدارس باستخدام كافة المعلومات المتوفرة من خلال عينة مدروسة للوصول إلى عدد واحد أو نقطة تكون تقديراً للوسيط المراد تقديره . فمثلا سنجد مستقبلا أن وسط العينة $\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$ يمثل تقديراً نقطياً لوسط المجتمع μ .

(٦,٢) طرق التقدير الكلاسيكية Classical estimation methods

إن تقدير وسيط فى مجتمع إحصائى مدروس يمكن أن يعطى كتقدير نقطى أو كتقدير مجالى . يرمز عادة للتقدير النقطى للوسيط θ فى مجتمع إحصائى بالرمز $\hat{\Theta}$ ولقيمته بالرمز $\hat{\theta}$. فمثلا تمثل القيمة $\bar{x}$ للإحصاء $\bar{X}$ والمحسوبة من خلال عينة حجمها n تقديراً نقطياً للوسط μ فى أى مجتمع إحصائى مدروس .

يسمى الإحصاء المستخدم للحصول على تقدير نقطى بتقدير (estimator) أو بدالة القرار (decision function) . لذلك فإن دالة القرار S التابعة لقيم العينة تمثل تقديرا للوسيط σ ، والتقدير S هو الإجراء المأخوذ من عينات مختلفة والذى سيؤدى بشكل عام إلى إجراءات مختلفة أو تقديرات .

تعريف (٦,١) فضاء القرار Decision space

إن مجموعة كل الإجراءات الممكنة والمأخوذة فى تقدير وسيط ، تدعى الإجراءات أو فضاء القرار .

ولا يمكن تقدير وسيط لمجتمع إحصائى بدون ارتكاب خطأ . فلا نتوقع مثلا أن تكون $\bar{X}$ تقديرا للوسيط μ بدون أخطاء . ونأمل ألا يكون هذا التقدير بعيدا جدا عن القيمة الحقيقية

للوسيط μ . فمن الممكن الحصول على تقدير قريب جدا من الوسيط المجهول μ بالنسبة لعينة خاصة وذلك باستخدام متوسط هذه العينة $\tilde{X}$ كتقدير μ . لنعتبر مثلا أن عينة خاصة تتألف من النقاط 4, 8, 9 مأخوذة من مجتمع وسطه $\mu = 6$ (ولكن نفترض أنه مجهول) . نستطيع أن

نقدر μ بواسطة العينة $\bar{x} = 7$ أو $\tilde{x} = 8$ باستخدام متوسط العينة كتقدير لهذا الوسيط . وفى هذه الحالة نلاحظ أن متوسط العينة $\tilde{X}$ يمثل تقديراً أقرب إلى القيمة الحقيقية للوسط μ من التقدير $\bar{X}$ الممثل لوسط العينة . من ناحية أخرى إذا احتوت عينتا على القيم 4, 7, 8, 5 ، فعندئذ نجد أن $\tilde{x} = 7$ و $\bar{x} = 6.5$ ، ولذلك فإن وسط العينة $\bar{X}$ يمثل تقديرا أفضل فى هذه الحالة . وفى حال عدم معرفة القيمة الحقيقية للوسط μ ، علينا أن نحدد سلفاً أيهما سيكون أفضل $\bar{X}$ أو $\tilde{X}$ لاستخدامه كتقدير للوسط μ المجهول .

وما يخطر على بال الدارس هو معرفة الخواص التى يجب أن تتمتع بها دالة القرار والتى تؤثر علينا فى اختبار تقدير دون آخر مختلف عنه .

لنفرض أن $\hat{\Theta}$ تقديرا تمثل قيمته $\hat{\theta}$ تقديرا نقطيا للوسيط المجهول θ فى مجتمع إحصائى .

تعريف (٦,٢) تقدير غير متحيز Unbiasel - estimator

نسمى الإحصاء $\hat{\Theta}$ تقديراً غير متحيز (unbiased-estimator) للوسيط θ المجهول إذا تحقق الشرط التالى :

$$E[\hat{\Theta}] = \theta$$

مثال (٦,١)

نلاحظ أن $\bar{X}$ هو تقدير غير متحيز للوسط μ لأى مجتمع . ذلك لأن :

$$E(\bar{X}) = E\left(\frac{\sum_{i=1}^{n} X_i}{n}\right) = \frac{\sum_{i=1}^{n} E(X_i)}{n} = E(X_1) = \mu$$

مثال (٦,٢)

بين أن S^2 يمثل تقديراً غير متحيز للتباين σ^2 .

الحل

$$\sum_{i=1}^{n} (X_i - \overline{X})^2 = \sum_{i=1}^{n} [(X_i - \mu) - (\overline{X} - \mu)]^2$$

نلاحظ أن

$$= \sum_{i=1}^{n} (X_i - \mu)^2 - 2 (\overline{X} - \mu) \sum_{i=1}^{n} (X_i - \mu) + n (\overline{X} - \mu)^2$$

$$= \sum_{i=1}^{n} (X_i - \mu)^2 - n(\overline{X} - \mu)^2$$

ولكن :

$$E (S^2) = E \left[\frac{\sum^{n}(X_i - \overline{X})^2}{n - 1} \right]$$

$$= \frac{1}{n-1} \left[\sum_{i=1}^{n} E (X_i - \mu)^2 - n E(\overline{X} - \mu)^2 \right]$$

$$= \frac{1}{n-1} \left(\sum_{i=1}^{n} \sigma^2_{x_i} - n \sigma^2_{\overline{x}} \right)$$

غير أنه من أجل $i = 1, 2, \ldots, n$ لدينا :

$$\sigma^2_{x_i} = \sigma^2$$

ومنه :

$$\sigma^2_{\overline{x}} = \frac{\sigma^2}{n}$$

وأخيراً فإننا نستنتج أن :

$$E(S^2) = \frac{1}{(n - 1)} \left(n \sigma^2 - n \frac{\sigma^2}{n} \right) = \sigma^2$$

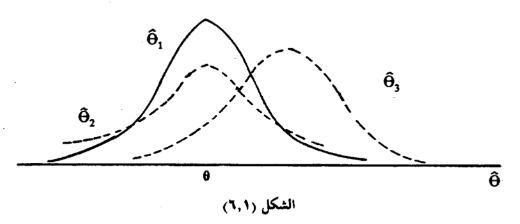
والعلاقة الأخيرة توضح أن S^2 يمثل تقديراً غير متحيز للوسيط σ^2 ، ومن جهة ثانية فإن S هو تقدير متحيز لـ σ بانحياز ضئيل وتافه من أجل عينات كبيرة الحجم .

إذا كان $\hat{\theta}_2 , \hat{\theta}_1$ يمثلان تقديرين غير متحيزين لنفس الوسيط θ فى مجتمع إحصائى ، فإننا نختار التقدير ذى التباين الأقل . لذلك إذا كان $\sigma^2_{\hat{\theta}_1} < \sigma^2_{\hat{\theta}_2}$ فإننا نقول بأن $\hat{\theta}_1$ أكثر فعالية من التقدير $\hat{\theta}_2$.

تعريف (٦,٣) التقدير غير المتحيز للوسيط Unbiased estimator for median

نقول بأن التقدير غير المتحيز $\hat{\theta}$ للوسيط θ أكثر فعالية إذا كان تباينه أقل من تباين أى تقدير غير متحيز لنفس الوسيط .

يوضح الشكل (٦,١) توزيعات المعاينة لثلاثة تقديرات $\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3$ للوسيط θ. ومن الواضح أن كلا من $\hat{\theta}_2, \hat{\theta}_1$ يمثل تقديرا غير متحيز للوسيط θ لأن توزيع كل من هذين التقديرين متناظراً بالنسبة لـ θ . كما أن للتقدير غير المتحيز $\hat{\theta}_1$ تباينا أقل من تباين $\hat{\theta}_2$ ، ولذلك فإن $\hat{\theta}_1$ أكثر فعالية من $\hat{\theta}_2$ ، وعلى هذا الأساس فإننا نختار من بين التقديرات الثلاثة السابقة كتقدير للوسيط θ التقدير $\hat{\theta}_1$.



الشكل (٦,١)

لقد أوضحنا أن $\bar{X}$ يمثل تقديرا غير متحيز للوسط μ فى أى مجتمع ، ويمكن بسهولة أن نرى أن $\tilde{X}$ متوسط العينة ما هو إلا تقدير غير متحيز للوسط μ فى المجتمع الطبيعى أيضا . غير أن تباين التقدير $\bar{X}$ أصغر من تباين التقدير $\tilde{X}$. وهكذا فإن كلا من القيمتين $\bar{x}$ و $\tilde{x}$ ستكون فى المتوسط مساوية لقيمة المجتمع μ ، ولكن $\bar{X}$ ستكون (من أجل عينة معينة) أقرب إلى μ .

إن المجال الذى يحتوى فيه التقدير يسمى بمجال التقدير . وهو مجال محدود العرض مركزه التقدير نفسه . ويجب أن يحتوى هذا المجال على القيمة الحقيقية للوسيط المجهول ،

وللحصول على مجال التقدير للوسيط المجهول Θ علينا أن نشكل مجالا من الشكل $\hat{\Theta} \pm k$ حيث تتعلق $\hat{\Theta}$ بالعينة الخاصة المختارة ويتحدد العدد k بواسطة توزيع المعاينة للاحصاء $\hat{\Theta}$. والآن فإن إدعاءنا بأن التقدير $\hat{\Theta}$ مساو تماما للوسيط Θ يعنى أن :

$$\hat{\Theta} - k < \Theta < \hat{\Theta} + k$$

يمكن بواسطة توزيع المعاينة للإحصاء θ أن نحدد قيمة k بحيث يكون الاحتمال $P[\hat{\theta} - k < \theta < \hat{\theta} + k]$ مساوياً لعدد معين يهم الدارس . فإذا حسبنا مثلا k بحيث يكون $P[\hat{\theta} - k < \theta < \hat{\theta} + k] = 0.95$ فعندئذ ستعطينا العينة المسحوبة عشوائياً مجالا يحتوى على الوسيط المجهول باحتمال قدره 0.95 . يسمى المجال السابق المحسوب من خلال عينة عشوائية بـ ±95 مجال ثقة . وهذا الاحتمال يعني أن ±95 من ثقتنا تؤكد بأن مجالنا هذا سيحتوى على القيمة الحقيقية للوسيط في المجتمع المدروس ، وكذلك الأمر إذا كان الاحتمال مساوياً لـ ±0.99 . فعندئذ ستكون قيمة k أكبر ، وبذلك يصبح عرض مجال الثقة أكبر . ونلاحظ أنه كلما عظمت الثقة كلما عرض المجال . وبشكل عام فإن توزيع الإحصاء $\hat{\theta}$ سيمكننا من حساب k بحيث يكون :

$$P[\hat{\theta} - k < \theta < \hat{\theta} + k] = 1 - \alpha \quad : \quad 0 < \alpha < 1$$

نسمي المجال السابق المحسوب من خلال عينة خاصة بـ $(1 - \alpha)100\%$ مجال ثقة . فمثلا من أجل $\alpha = 0.05$ نجد 95% مجال ثقة . ومن أجل $\alpha = 0.01$ نجد 99% مجال ثقة . تدعى النسبة $(1 - \alpha)$ معامل الثقة . وحدود المجال $\hat{\theta} - k$ و $\hat{\theta} + k$ بنهايات الثقة أو نهايات الإسناد .

(٦,٣) تقدير الوسط Estimating the mean

كتقدير للوسط μ في مجتمع إحصائي بأخذ عادة الإحصاء $\bar{X}$. ومن المعلوم أن توزيع المعاينة للإحصاء $\bar{X}$ يتمركز حول الوسط μ ، وفي أغلب التطبيقات يكون تباين $\bar{X}$ أصغر من تباين أى تقدير آخر لـ μ . ولهذا فإن وسط العينة $\bar{x}$ سيستخدم كتقدير نقطى للوسط μ في أى مجتمع إحصائي ، وإذا تذكرنا بأن $\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}$ أدركنا أن العينة الكبيرة ستقدم لنا قيمة لـ $\bar{X}$ ذات تباين صغير . لذلك فإن من المرجح أن تكون $\bar{X}$ قريبة جداً من μ عندما تكون n كبيرة . للبحث عن مجال ثقة للوسط μ نفرض أن العينة المدروسة مسحوبة من مجتمع طبيعي أو أنها حجم كبير وذلك بغض النظر عن نوع

المجتمع المدروس، ولنفرض أن تباين المجتمع σ^2 معلوم . يمكن أن تشكل مجال ثقة للوسط μ بالاعتماد على توزيع المعاينة لـ $\overline{X}$. وحسب نظرية النهايات المركزية يمكننا أن نتوقع أن يكون لتوزيع المعاينة $\overline{X}$ توزيعاً قريباً من التوزيع الطبيعي بالوسط $\mu = \mu_{\overline{x}}$ والانحراف

المعياري $\sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}}$. فإذا فرضنا أن $z_{\alpha/2}$ هي نقطة من محور السينات تحدد على يمينها مساحة قدرها $\alpha/2$ لأمكننا كتابة :

$$P(-Z_{\alpha/2} < Z < Z_{\alpha/2}) = 1 - \alpha$$

حيث إن :

$$Z = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}}$$

لذلك فإن :

$$P\left(-Z_{\alpha/2} < \frac{\overline{X} - \mu}{\sigma / \sqrt{n}} < Z_{\alpha/2}\right) = 1 - \alpha$$

لنضرب كل طرف من أطراف المتباينة السابقة بالعدد $\sigma/\sqrt{n}$ ، ثم لنطرح من جميع الأطراف المقدار $\overline{X}$ فنجد بعد أن نضرب الأطراف بإشارة ناقص أن :

$$P\left(\overline{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

وهكذا نجد أن $(1 - \alpha)100\%$ مجال ثقة للوسط μ في مجتمع فيه σ معلومة هو المجال :

$$\left(\overline{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} , \overline{X} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$$

ملاحظة

إذا كانت العينة مسحوبة من مجتمع غير طبيعي وكان حجمها صغيراً ، فإنه لا

يمكننا التوقع بأن درجة الثقة مضبوطة (أو دقيقة) ، ومع ذلك فإن نظرية العينات تضمن لنا نتيجة جيدة من أجل عينات حجمها $n \geq 30$ بغض النظر عن شكل المجتمع المدروس .

للبحث عن $(1 - \alpha)100\%$ مجال ثقة للوسط μ نفترض أن σ معلوم . وبما أن هذه الحالة لا تتحقق بشكل دائم ، لذلك يمكن أن نستبدل σ بـ S شريطة أن تكون $n \geq 30$.

مثال (٦,٣)

بفرض أن أطوال خمسين طالباً (مسحوبين بشكل عشوائي من مجتمع ما من الطلبة) قد قدم لنا وسطاً قدره $\bar{x} = 174.5$ cm وانحرافاً معيارياً قدره $S = 6.9$ cm . لننشيء 98% مجال ثقة للوسط μ (أي متوسط طول كل طالب من طلاب المجتمع المدروس) نلاحظ أن التقدير النقطي للوسط μ (الممثل لمتوسط طول كل طالب) ما هو إلا $\bar{x} = 174.5$ cm ، وبما أن حجم العينة المسحوبة من الطلبة هو $n = 50$ ، لذلك يمكن أن نقرب الانحراف المعياري σ بواسطة $S = 6.9$ cm . النقطة Z من محور السينات والتي تحدد على يمينها مساحة قدرها 0.01 أو تحدد على يسارها مساحة قدرها 0.99 هي النقطة $Z_{\alpha/2}$ (أنظر الجدول IV) . لذلك فإن 98% مجال الثقة هو المجال :

$$\left(174.5 - (2.33) \left(\frac{6.9}{\sqrt{50}} \right) 174.5 + (2.33) \left(\frac{6.9}{\sqrt{50}} \right) \right) = (172.23 \ , \ 176.77)$$

وهكذا نجد أن : $172.23 < \mu < 176.77$

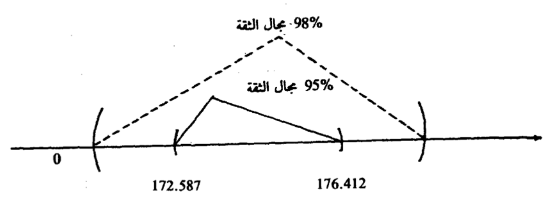
لنبحث عن 95% مجال ثقة لنفس الوسط μ . إن النقطة Z من محور السينات والتي تحدد على يمينها مساحة قدرها 0.025 أو إنها تحدد على يسارها مساحة قدرها 0.975 هي النقطة $Z_{\frac{\alpha}{2}} = 1.9$ ، ولذلك فإن 95% مجال ثقة للوسط μ ما هو إلا المجال :

$$\left(174.5 - (1.96) \left(\frac{6.9}{\sqrt{50}} \right) , 174.5 + (1.96) \left(\frac{6.9}{\sqrt{50}} \right) \right)$$

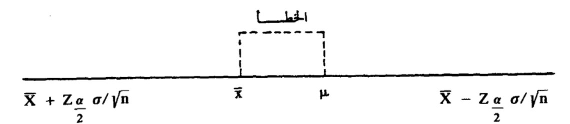
أو : $(172.587 \ , \ 176.412)$

أي إن : $172.587 < \mu < 176.412$

ويوضح الشكل التالى طولى مجالى الثقة :



من الشكل السابق نستنتج أن مجال الثقة للأطوال يتطلب تقديرا لـ μ بدرجة أكبر من الدقة ويزودنا $100\%\ (1 - \alpha)$ مجال الثقة بتقدير درجة دقة تقديرنا النقطى ، فإذا كان الوسط المجهول μ فعلياً هو مركز المجال فإننا نجد أن التقدير $\bar{X}$ يقدر μ بدون خطأ . وعلى الغالب لا يكون $\bar{X}$ مساويا تماماً لـ μ ، ولذلك فإن $\bar{X}$ يقدر μ بخطأ . وحجم الخطأ هذا ما هو إلا الفرق بين القيمتين μ و $\bar{X}$. ويمكننا أن نكون واثقين بنسبة $(1 - \alpha)$ 100% من أن هذا الفرق سيكون أقل من $Z_{\alpha/2}\, \sigma/\sqrt{n}$ ، ويمكننا بسهولة من التحقق مما سبق برسم مخطط لمجال الثقة كما هو موضح على الشكل (٦,٢)



الشكل (٦,٢) يوضح الخطأ المرتكب فى تقدير μ بـ $\bar{X}$

نظرية (٦,١)

إذا كان $\bar{x}$ تقديرا للوسط μ فى مجتمع احصائى ما . فإنه يمكننا أن نثق بنسبة قدرها $100\%(1 - \alpha)$ من الخطأ سيكون أقل من $Z_{\alpha/2}\sigma/\sqrt{n}$. ففى المثال (٦,٣) نجد أنه بنسبة قدرها 98% يمكن أن نثق بأن وسط العينة $\bar{x} = 174.5$ cm سيختلف عن القيمة

الحقيقية للوسط μ (الممثل لطول كل طالب) بمقدار أقل من 2.273 cm ، وأن نثق بنسبة قدرها 95% من أن الفرق سيكون أقل من 1.91 cm .

كثيرا ما نهتم فى معرفة حجم العينة اللازم للتأكد من أن الخطأ المرتكب فى تقدير μ سيكون أقل من عدد معين e بحسب النظرية (٦,١) نرى أن علينا أن نختار n بحيث يكون : $Z_{\alpha/2}\, \sigma/\sqrt{n} = e$.

$$Z_{\frac{\alpha}{2}}\, \sigma/\sqrt{n} = e$$

نظرية (٦,٢)

إذا استخدم الإحصاء $\bar{X}$ كتقدير للوسط μ فى مجتمع إحصائى ما ، فيمكننا أن نثق بنسبة $(1-\alpha)100\%$ من أن الخطأ المرتكب فى التقدير سيكون أقل من عدد معين e عندما يكون حجم العينة المختارة مساويا لـ :

$$n = \left(\frac{Z_{\alpha/2}\, \sigma}{e} \right)^2$$

وتكون النظرية (٦,٢) السابقة جاهزة للتطبيق فيما إذا كنا نعرف تباين المجتمع الإحصائى الذى نختار منه عينتا . وفى الحالة التى تكون فيها σ مجهولة ، فإن بإمكاننا أن نأخذ عينة (تمهيدية) ذات حجم $n \geq 30$ وذلك بغية الحصول على تقدير لـ σ ، وبعد الحصول على هذا التقدير نستخدم النظرية (٦,٢) وبواسطتها نستطيع تقريبا أن نحدد عدد الملاحظات التى نحتاجها للحصول على الدرجة المطلوبة من الدقة .

مثال (٦,٤)

ما هو حجم العينة المطلوب فى المثال (٦,٣) إذا أردنا أن يكون الخطأ المرتكب فى تقدير μ أقل من 2.273 باحتمال ثقة قدره 98% ؟

الحل

من الواضح فى المثال (٦,٣) المحسوب من عينة تمهيدية حجمها $n = 50$ أن $S = 6.9$ cm ، لهذا يمكننا استخدامه عوضا عن σ . وبحسب النظرية (٦,٢) نجد أن :

$$n = \left[\frac{(2.33)\,(6.9)}{2.273} \right]^2 = 50.02$$

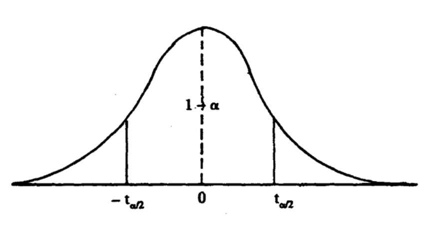
نستنتج مما سبق أن العينة المختارة للتقدير والتى حجمها $n = 50$ ستقدم لنا تقديرا لـ μ لا يتجاوز الخطأ المرتكب فيه 2.273 بثقة قدرها 98% .

غالبا ما نفكر فى إيجاد تقدير للوسط μ فى مجتمع إحصائى تباينه σ^2 مجهول . ولكن غالبا ما تضع كلفة العينة والوقت المتوفر ، بالإضافة إلى عوامل أخرى ، حدا لحجم العينة ، بحيث يصبح الحجم الكبير المطلوب غير عملى ، وفى هذه الحالة لا يمكن اللجوء إلى طرق الاستقراء الموافق لعينات حجمها $n \geq 30$.

ولقد ذكرنا سابقا أن للكمية $\frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$ توزيعاً طبيعياً معياريا إذا كان المجتمع المدروس طبيعياً ، أو توزيعاً قريباً من التوزيع الطبيعى المعيارى إذا كان المجتمع المدروس غير طبيعى ، وذلك طبقا لنظرية النهاية المركزية بشرط أن يكون حجم العينة n كبيراً . ولكن فى الحالة التى تكون فيها n صغيرة ، وحيث أنه لا يجوز لنا اللجوء الى نظرية النهاية المركزية ، فمن الطبيعى وضع S محسوبة من العينة (التمهيدية) بدلا من σ . وبما أن العينة صغيرة ، فإن هذا يجعلنا نشك فى أن تكون S ممثلة لقيمة تقريبية . وهكذا نفقد كل أساس نظرى أو منطقى يسمح لنا بالقول بأن توزيع المتغير الجديد $\frac{\bar{X} - \mu}{S / \sqrt{n}}$ هو التوزيع الطبيعى المعيارى . ومع أن توقع S^2 هو σ^2 $(E(S^2) = \sigma^2)$ تماماً إلا أن احتمال كون S^2 أقل من σ^2 يزيد على النصف كما نرى من توزيع S^2 [يجب ألا يغيب على ذهن الطالب من أن $\frac{(n-1)\,S^2}{\sigma^2}$ له توزيع $\chi^2_{(n-1)}$] وهذا يحملنا على الاعتقاد بأن منحنى الكثافة للمتحول الجديد $\frac{\bar{X} - \mu}{S / \sqrt{n}}$ سيكون أكثر انتشارا .

وعندما تزداد n ، فإن S^2 تقترب من σ^2 ويقترب التوزيع الجديد من التوزيع الطبيعى المعيارى . ويبرهن فى الإحصاء الرياضى أنه إذا كانت $X_1, X_2 \ldots X_n$ عينة من مجتمع طبيعى وسطه μ وتباينه σ^2 ، فعندئذ يكون للمتغير $t = \frac{\bar{X} - \mu}{S / \sqrt{n}}$ توزيعا عينيا هو التوزيع t بـ $(n - 1)$ درجة حرية والذى سبق ذكره فى الفصل الخامس .

يوضح الشكل (٦,٣) توزيع المتغير الجديد $t = \frac{\bar{X} - \mu}{S / \sqrt{n}}$



الشكل (٦,٣)

من هذا الشكل نجد أن :

$$P(-t_{\alpha/2} < T < t_{\alpha/2}) = 1 - \alpha$$

حيث ترمز $t_{\alpha/2}$ إلى قيمة t التى تحصر على يمينها مساحة قدرها $\alpha/2$ ، ومن تناظر الشكل نجد أن مساحة قدرها $\alpha/2$ ستقع على يسار النقطة $t = -t_{\alpha/2}$. وهكذا نجد أن :

$$P\left(-t_{\alpha/2} < \frac{\overline{X} - \mu}{S/\sqrt{n}} < t_{\alpha/2}\right) = 1 - \alpha$$

ويضرب كل طرف من أطراف المتباينة السابقة بـ $S/\sqrt{n}$ ثم بطرح $\overline{X}$ وضرب جميع الأطراف بـ 1 - نجد أن :

$$P\left(\overline{X} - \frac{t_{\alpha/2} \cdot S}{\sqrt{n}} < \mu < \overline{X} + t_{\alpha/2}\frac{S}{\sqrt{n}}\right) = 1 - \alpha .$$

وهكذا نستنتج $(1-\alpha)100\%$ مجال ثقة للوسط μ فى مجتمع تباينه σ^2 مجهول ، وذلك من أجل عينات حجمها $n < 30$.

مثال (٦,٥)

يزيد وزن نوع معين من الطيور بمقدار 65 غراما خلال الأشهر الثلاثة الأولى من حياتها . وقد أطعمت 12 من هذه الطيور وفقا لنظام تغذية معين خلال الأشهر الثلاثة الأولى وقيست الزيادة فى وزن كل منا فوجدت :

55, 62, 54, 58, 65, 64, 60, 60, 62, 59, 67, 62, 61

فالمطلوب البحث عن 95% مجال ثقة للوسط μ الممثل لمتوسط الزيادة فى الوزن خلال الأشهر الثلاثة الأولى .

الحل

نلاحظ أن $n = 12$ وأن $n - 1 = 11$.

ثم إن :

$$\overline{x}_0 = \frac{\sum_{i=1}^{12} x_i}{12} = 60.75$$

$$S^2 = \frac{\sum_{i=1}^{12} (x_i - \overline{x}_0)^2}{(n-1)}$$

$$S = 3.8406$$

نلاحظ أن :

$$\overline{X} = 60.75$$

ومجال الثقة المطلوب هو :

$$\left(\overline{X} - \frac{t_{/2} \cdot S}{\sqrt{n}}, \overline{X} + \frac{t_{\alpha/2} \cdot S}{\sqrt{n}} \right)$$

ومن الجدول VI وبفرض أن $n - 1 = 11$ نجد أن :

$$t_{\alpha/2}(n-1) = t_{0.025}(11) = 2.20$$

وهكذا نجد أن مجال الثقة هو :

$$60.75 - (2.20)\frac{(3.8406)}{\sqrt{12}} , 60.75 + (2.20)\frac{(3.8406)}{\sqrt{12}}$$

وأخيراً فإن 95% مجال ثقة للوسط المجهول μ هو المجال :

(58.3108, 63.1891)

(٦,٤) تقدير فرق وسطين Estimating the difference between two means

توازى مسألة تقدير الفرق بين وسطين فى أهميتها مسألة تقدير وسط مجتمع إحصائى . فقد يرغب أحدنا فى مقارنة طريقتين فى التعليم ، فنقسم الطلبة عشوائيا إلى زمرتين ، ونُخضع كلا منهما لطريقة تعليم معينة ، ثم نقوم بالاستقراء حول الفرق بين تحصيل الطلبة فى كل من الزمرتين وذلك بواسطة مقياس معين . أو قد نرغب فى مقارنة معدلى الإنتاج فى شكرة للصناعة النسيجية عند استخدام المواد الأولية الواردة من مموليْن B, A ، فنأخذ عينة من الإنتاج اليومى فى كل حالة ونستخدم المعلومات التى تقدمها هاتين العينتين للقيام باستقراء يتعلق بالفرق بين معدلى الإنتاج .

فبفرض أن المجتمع الإحصائى الأول بالوسط μ_1 والتباين σ_1^2 ، والثانى بالوسط μ_2 والتباين σ_2^2 ، وأننا سحبنا عينة عشوائية ذات حجم n_1 من المجتمع الأول ، وعينة أخرى ذات حجم n_2 من المجتمع الثانى ولنفرض أن العينتين قد سحبتا بحيث تكون كلا منهما مستقلة عن الأخرى . بعد ذلك نقوم بحساب تقديرات لوسطاء المجتمعين وهى $\overline{X}_1$, S_1^2 من العينة الأولى و $\overline{X}_2$, S_2^2 من العينة الثانية ، والتقدير النقطى للفرق بين وسطى المجتمعين $\mu_1 - \mu_2$ هو $\overline{X}_1 - \overline{X}_2$. وكما نعلم من خواص التوقع أن :

$$E(\overline{X}_1 - \overline{X}_2) = E(\overline{X}_1) + E(-\overline{X}_2) = E(\overline{X}_1) - E(-E(\overline{X}_2) = \mu_1 - \mu_2$$

وبما أن العينتين مستقلتان لذلك فإنه يمكننا كتابة :

$$\sigma^2_{\overline{X}_1 - \overline{X}_2} = \sigma^2_{\overline{X}_1} + \sigma^2_{\overline{X}_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

ومنه :

$$\sigma^2_{\overline{X}_1 - \overline{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

وبالاضافة لذلك إذا كانت كل من n_2 , n_1 أكبر من تساوى 30 ، فعندئذ حسب النظرية (٥,١٥) سيكون للمتغير :

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

توزيعاً قريباً من التوزيع الطبيعى المعيارى .

أى إنه تمكن أن نجزم باحتمال قدره $(1 - \alpha)$ من أن هذا المتغير الطبيعى المعيارى سيقع ضمن العددين $Z_{\alpha/2}$ و $Z_{\alpha/2}$ – أى أن :

$$P [- Z_{\alpha/2} < Z < Z_{\alpha/2}] = 1 - \alpha$$

وبتعويض قيمة Z فى المعادلة السابقة نجد أن :

$$P \left[-Z_{\alpha/2} < \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} < Z_{\frac{\alpha}{2}} \right] = 1 - \alpha$$

وبضرب كل حد من حدود المتباينة السابقة بالعدد الموجب :

$$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

ثم بطرح العدد $(\bar{X}_1 - \bar{X}_2)$ من كل حد من حدودها ، وأخيراً بضرب جميع حدود المتباينة بـ 1 – نجد أن :

$$P \left[(\bar{X}_1 - \bar{X}_2) - Z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_2 - \mu_2 < (\bar{X}_1 - \bar{X}_2) + Z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right] = 1 - \alpha$$

وهكذا نصل إلى التعريف التالى :

تعريف مجال ثقة فرق وسطين $\mu_1 - \mu_2$

إن $(1 - \alpha)100\%$ مجال ثقة فرق وسطين $\mu_1 - \mu_2$ هو المجال :

$$\bar{X}_1 - \bar{X}_2 - Z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{X}_1 - \bar{X}_2) + Z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

وذلك بفرض أن تبايني المجتمعين المدروسين σ_1^2 , σ_2^2 معلومان ، وأن $Z_{\alpha/2}$ هي النقطة من المحور السينى تحت المنحنى الطبيعى المعيارى والتى تحدد تحت المنحنى ، وعلى يمينها مساحة قدرها $\alpha/2$.

ملاحظة :

كما ذكرنا سابقا إذا كان σ_1^2 , σ_2^2 مجهولين وكانت العينات المختارة كبيرة بشكل كاف (أى أن $n_2 \geq 30$ و $n_1 \geq 30$) فعندئذ يمكن استبدال σ_1^2 , σ_2^2 بـ S_1^2 , S_2^2 على التتالي .

مثال (٦,٦)

لدى مقارنة نوعين من إطارات السيارات باختبار عملى أخذنا من كل نوع عينة حجمها $n = 100$ إطاراً ، ثم سجلنا عدد الأميال التى خدمها كل إطار حتى اهترائه وفقا لمقاييس محددة سلفا . فإذا كانت نتائج الاختبار بالأميال كالتالى :

$$\bar{X}_1 = 26400 \qquad S_1^2 = 1440000$$
$$\bar{X}_2 = 25100 \qquad S_2^2 = 1960000$$

فما هو تقدير الفرق بين وسطي العمر فى النوعين ، ثم ما ماهو حدود الخطأ فى التقدير ؟

الحل

نلاحظ أن التقدير النقطى لـ $(\mu_1 - \mu_2)$ [حيث يمثل μ_1 العمر الوسطى لإطار من النوع الأول ، μ_2 العمر الوسطى للإطار من النوع الثانى] هو $\overline{X}_1 - \overline{X}_2$ أى :

$$\overline{X}_1 - \overline{X}_2 = 26400 - 25100 = 1300$$

وكذلك نجد أن :

$$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} = 184$$

ونلاحظ أن حدود الخطأ هو حوالى $2\sigma^2_{\bar{x}_2 - \bar{x}_2} = 368$ ميلاً ، وهكذا نجد أن النوع الأول متفوق على النوع الثانى .

مثال (٦,٧)

لنبحث عن 0.99 مجال ثقة للفرق $\mu_1 - \mu_2$ فى المثال (٦,٦) .

الحل

نعلم أن $Z_{\alpha/2} = 2.58$ ومنه فمجال الثقة المطلوب هو

$$(\bar{x}_1 - \bar{x}_2) - 2.58\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\ ,\ (\bar{x}_1 - \bar{x}_2) + 2.58\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

وباستخدام نتائج التمرين (٦,٦) نجد أن 0.99 مجال ثقة لفرق الوسطين μ_1 , μ_2 هو المجال :

$$(1300 - 2.58\ (184)\ ,\ 1300 + 2.58\ (184))$$

أو المجال :

$$(825.28\ ,\ 1774.72)$$

أى أن الحد الأدنى للثقة هو 825 ، والحد الأعلى للثقة هو 1775 ، أى أنه يمكن تقدير الفرق بين وسطي العمرين بأنه يقع بين هذين الحدين .

ملاحظة

إذا كانت العينات المختارة من المجتمعين صغيرة الحجم ، فعندئذ نلجأ إلى التوزيع t لإيجاد مجالات ثقة وهذا الأمر صحيح وشرعى عندما تكون المجتمعات المدروسة لها تقريبا توزيعات طبيعية .

لنفرض الآن أن σ_1^2 , σ_2^2 مجهولين وأن n_1 , n_2 صغيرين ، وكلا منهما أقل من 30 .
إذا كان $\sigma_1^2 = \sigma_2^2 = \sigma^2$ فإننا نحصل على متغير عشوائى طبيعى معيارى من الشكل :

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\sigma^2 \left[\frac{1}{n_1} + \frac{1}{n_2} \right]}}$$

وبحسب النظرية (٥,١٦) يكون للمتغيرين $(n_1 - 1)S_1^2/\sigma^2$, $(n_2 - 1)S_2^2/\sigma^2$ توزيعين من نوع χ^2 (كاى – مربع) بعدد من درجات الحرية مساو على الترتيب $(n_1 - 1)$, $(n_2 - 1)$.
وهذين المتغيرين مستقلين أيضا (لأن العينات المختارة مستقلة) . لذلك فإن مجموعهما :

$$V = \frac{(n_1 - 1) \ S_1^2}{\sigma^2} + \frac{(n_2 - 1)S_2^2}{\sigma^2} \quad = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{\sigma^2}$$

له توزيع من نوع χ^2 (كاى – مربع) بعدد من درجات الحرية مساو لـ $(n_1 + n_2 - 2)$.
وبتعويض V , Z فى النظرية (٥,١٧) فإننا نحصل على الإحصاء التالى :

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\sigma^2 \left[\frac{1}{n_1} + \frac{1}{n_2} \right]}} \sqrt{\frac{(n_1 - 1) S_1^2 + (n_2 -) S_2^2}{\sigma^2 (n_1 + n_2 - 2)}}$$

ويتوزع هذا المتغير وفقا للتوزيع t بعدد من درجات الحرية مساو لـ $(n_1 + n_2 - 2)$ ، وأن التقدير النقطى للتباين المجهول σ^2 يمكن الحصول عليه بواسطة المجموع :

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

وبتبديل S_p^2 فى عبارة الإحصاء T نجد أن :

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

وباستخدام الإحصاء السابق نجد أن :

$$P[-t_{\alpha/2} < T < t_{\alpha/2}] = 1 - \alpha$$

حيث تمثل $t_{\alpha/2}$ القيمة t الواقعة على محور السينات أسفل منحنى التوزيع t بـ $(n_1 + n_2 - 2)$ درجة من الحرية ، والتى تحدد على يمينها مساحة قدرها $\alpha/2$. وبتبديل قيمة T فى المعادلة السابقة نجد أن :

$$P\left[-t_{\alpha/2} < \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} < t_{\alpha/2}\right] = 1 - \alpha$$

أو العلاقة :

$$P\left[(\bar{X}_1 - \bar{X}_2) - t_{\alpha/2} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} < \mu_1 - \mu_2 < (\bar{X}_1 - \bar{X}_2) + t_{\alpha/2} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right] = 1 - \alpha$$

فمن أجل عينتين عشوائيتين مستقلتين ذات حجمين n_1 , n_2 مختارتين بشكل عشوائى ، من مجتمعين طبيعيين نجد أن فرق وسطى العينتين $\bar{X}_1 - \bar{X}_2$ يمثل تقديرا نقطيا لفرق الوسطين $\mu_1 - \mu_2$ ، وأن $100\%(1 - \alpha)$ مجال الثقة لهذا التقدير هو المجال :

$$\left((\bar{X}_1 - \bar{X}_2) - t_{\alpha/2} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}, (\bar{X}_1 - \bar{X}_2) + t_{\alpha/2} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right)$$

حيث يمثل S_p الانحراف المعيارى المشترك للمجتمعين المدروسين .

مثال (٦,٨)

استخدم في عملية كيميائية عاملان مساعدان لمقارنة تأثيرهما على قدرة عملية

التفاعل . وقد تم تحضير اثنى عشر مزيجا باستخدام العامل المساعد الأول ، فأعطت هذه الأمزجة وسطا قدره $\bar{X}_1 = 85$ وانحرافا معياريا قدره $S_1 = 4$ ، كما تم تحضير عشرة أمزجة باستخدام العامل المساعد الثانى ، فأعطت وسطا قدره $\bar{X}_2 = 81$ وانحرافاً معيارياً قدره $S_2 = 5$.

فالمطلوب البحث عن 90% مجال ثقة للفرق بين وسطى المجتمعين ، وذلك بفرض أن للمجتمعين توزيعاً قريباً من التوزيع الطبيعى بنفس التباين σ^2 . بفرض أن μ_1 , μ_2 يمثلان على الترتيب وسطى المجتمعين اللذين أخذ منهما العاملان المساعدان الأول والثانى .

فعندئذ يكون المطلوب البحث عن 90% مجال ثقة للفرق $\mu_1 - \mu_2$.

نأخذ $\bar{X}_1 - \bar{X}_2 = 85 - 81 = 4$ كتقدير نقطى للفرق فنلاحظ أن :

$$S_p^2 = \frac{(n-1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

$$= \frac{(11)(16) + (9)(25)}{12 + 10 - 2} = 20.05$$

وبأخذ الجذر التربيعى للطرفين نجد أن $S_p = 4.478$. والملاحظ أيضا أن $1 - \alpha = 0.90$ ، إذا $\alpha = 0.1$. ومن الجدول V نجد أن $t_{0.05} = 1.725$ من أجل عدد من درجات الحرية $\nu = n_1 + n_2 - 2 = 20$ درجة .

لذلك بالتعويض فى العلاقة :

$$(\bar{x}_1 - \bar{x}_2) - t_{\alpha/2} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) - t_{\alpha/2} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

نجد أن 90% مجال الثقة هو المجال :

$$4 - (1.725)(4.478)\sqrt{\frac{1}{12} + \frac{1}{10}} < \mu_1 - \mu_2 < 4 + (1.725)(4.478)\sqrt{\frac{1}{12} + \frac{1}{10}}$$

أى مجال :

$$0.69 < \mu_1 - \mu_2 < 7.31$$

وحيث أن نهايات الثقة موجبة ، لذلك فإننا نستنتج أن العامل المساعد الأول متفوق على العامل المساعد الثانى .

لنفتش عن مجال ثقة للفرق $\mu_1 - \mu_2$ من أجل عينات صغيرة ، وذلك بفرض أن تباينات المجتمعات المدروسة مجهولة ، وغير متساوية ، وحجوم العينات المدروسة مختلفة أيضا . فى هذه الحالة نلاحظ أن للإحصاء .

$$\hat{T} = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 , \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

توزيعاً قريباً من التوزيع t بعدد من درجات الحرية قدره :

$$= \frac{[\,(S_1^2 / n_1) + (S_2^2 / n_2)\,]^2}{[\,(S_1^2 / n_1)^2 / (n_1 - 1)\,] + [\,(S_2^2 / n_2)^2 / (n_2 - 1)\,]}$$

وحيث إن ν عدداً نادراً ، لذلك نقربه إلى أقرب عدد تام . وباستخدام الإحصاء $\hat{T}$ نجد أن :

$$P\,(-\,t_{\alpha/2} < T < t_{\alpha/2}) \approx 1 - \alpha$$

حيث يمثل $t_{\alpha/2}$ قيمة المتغير t بعدد من درجات الحرية مساو لـ ، والتى تحدد على يمينها مساحة قدرها $\alpha/2$ ، وبتعويض قيمة T نجد أن :

$$(\overline{X}_1 - \overline{X}_2) - t_{\alpha/2} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} < \mu_1 - \mu_2 < (\overline{X}_1 - \overline{X}_2) + t_{\alpha/2} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

حيث يمثل $\overline{X}_2$, $\overline{X}_1$ وسطى العينتين الصغيرتين المستقلتين ذات الحجمين n_2 , n_1 ، والمأخوذتين من مجتمعين لهما تقريباً توزيعاً طبيعياً . وكذلك فإن S_2^2 , S_1^2 يمثلان أيضا التباينين العينين لهاتين العينتين .

مثال (٦,٩)

لدى الرجوع إلى سجلات الأرصاد الجوية فى المملكة العربية السعودية ، وجد أن وسط هطول الأمطار التى هطلت خلال الخمسة عشر عاما الماضية على منطقة القصيم فى شهر نيسان هو 4.93 سنتيمتر ، وأن انحرافها المعيارى هو 1.14 سنتيمتر ، كما وجد أن وسط هطول الأمطار التى سقطت فى منطقة مجاورة للقصيم فى نفس الشهر خلال العشرة أعوام السابقة من نفس الشهر هو 2.64 سنتيمتر ، والانحراف المعيارى لها هو 0.66 سنتيمتر . لننشىء 95% مجال ثقة لفرق الوسطين الحقيقين لهطول الأمطار فى هاتين المنطقتين ، وذلك بفرض أن الملاحظات السابقة أخذت من مجتمعات طبيعية بتباينات مختلفة . من أجل منطقة القصيم لدينا $n_1 = 15, S_1 = 1.14, \bar{X}_1 = 4.93$ ، ومن أجل المنطقة المجاورة للقصيم لدينا $n_1 = 10, S_1 = 0.66, \bar{X}_2 = 2.64$. لنفرض أن μ_1 هو الوسط الحقيقى لهطول الأمطار فى القصيم خلال عام و μ_2 فى المنطقة المجاورة ، وحيث أن تباينى المجتمعين المدروسين الطبيعيين مختلفان وحجمى العينتين مختلفان أيضا ، لذلك يمكننا أن نجد 95% مجال ثقة تقريبياً بالاعتماد على التوزيع t بعدد من درجات الحرية :

$$\nu = \frac{[\,S_1^2/n_1 + S_2^2/n_2\,]^2}{[\,(S_1^2/n_1)^2/n_1 - 1\,] + [\,(S_2^2/n_1)^2/n_1 - 1\,]}$$

$$= \frac{[\,(1.14)^2/15 + (0.66)^2/10\,]^2}{\{\,[\,(1.14)^2/15]^2/14\,\} + \{\,(0.66)^2/10\,]^2/9\,\}}$$

ومنه :

$$\nu = 22.7 \approx 23$$

كما نلاحظ أن التقدير النقطى للفرق $\mu_1 - \mu_2$ هو $\bar{X}_1 - \bar{X}_2 = 4.93 - 2.64 = 2.29$ ، وحيث أن $1 - \alpha = 0.95$ ، لذلك فإن $\alpha = 0.05$. من الجدول V نجد أن $t_{\alpha/2} = t_{0.025} = 2.069$ من أجل $\nu = 23$. وبالتعويض فى العلاقة :

$$(\bar{x}_1 - \bar{x}_2) - t_{\alpha/2}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + t_{\alpha/2}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

نحصل على مجال الثقة :

$$2.02 < \mu_1 - \mu_2 < 2.56$$

لذلك فإن 95% من ثقتنا تؤكد أن المجال (2.02, 2.56) يحوى الفرق الحقيقى لمتوسطى هطول الأمطار فى هاتين المنطقتين .

(٦,٥) تقدير P في المجتمع الحداني Estimating P in bionomial proportions

إن أفضل تقدير نقطى لـ P فى المجتمع الحدانى ، هو وسط عدد النجاحات خلال n تكراراً . أى أن التقدير P هو :

$$\hat{P} = \frac{X}{n}$$

حيث يمثل X عدد النجاحات و n عدد التكرارات ، ونقصد بعبارة (أفضل تقدير) أن التقدير $\hat{P}$ غير متحيز وله تباين أصغر من تباين أى تقدير غير متحيز لـ $\hat{P}$.

وفقا لنظرية النهايات المركزية ، فإن التوزيع التقريبى لـ X هو التوزيع الطبيعى بالوسط np والتباين np.q وذلك من أجل قيمة كبيرة لـ n ، وبما أن 1/n يمثل عدداً ثابتاً لذلك ، حسب خاصة التوزيع الطبيعى المذكورة فى الفصل الخامس ، ويكون التوزيع التقريبى لـ $\hat{P} = \frac{1}{n}(X)$ هو أيضا التوزيع الطبيعى بالوسط :

$$E(\hat{P}) = E\left(\frac{X}{n}\right) = \frac{1}{n} \cdot E(X) = \frac{1}{n} \cdot np = p$$

والتباين :

$$\sigma^2_{\hat{p}} = \sigma^2_{\frac{X}{n}} = \frac{1}{n^2}\sigma^2_x = \frac{npq}{n^2} = \frac{pq}{n}$$

ويمكننا التأكد من أن :

$$P\left(-Z_{\alpha/2} < \frac{\hat{P} - p}{\sqrt{pq/n}} < Z_{\alpha/2}\right) = 1 - \alpha$$

حيث تمثل $Z_{\alpha/2}$ قيمة على المحور السينى تحت منحنى دالة الكثافة للمتغير الطبيعى المعيارى ، والتى تحدد على يمينها مساحة قدرها $\alpha/2$. هذا ويمكن كتابة العلاقة الأخيرة على النحو التالى :

$$P\left(\hat{P} - Z_{\alpha/2}\frac{\sqrt{Pq}}{n} < P < \hat{P} + Z_{\alpha/2}\frac{\sqrt{Pq}}{n}\right) = 1 - \alpha$$

والصعوبة التي تواجهنا هنا هي في حساب $\sigma_n \frac{Pq}{n}$ التي تعتمد على P وعلى $q = 1 - p$ وهما مجهولان . فإذا وضعنا $\hat{P}$, $\hat{q}$ بدلا عن P , q في عبارة الانحراف المعياري $\sqrt{\frac{pq}{n}}$ فإن الخطأ المرتكب سيكون صغيراً جداً من أجل قيمة كبيرة لـ n . وفي الحقيقة يتغير الانحراف المعياري ببطء شديد عندما تتحول P . ويمكن ملاحظة ذلك بوضوح في الجدول التالي :

P	$\sqrt{Pq}$
0.5	0.50
0.4	0.49
0.3	0.46
2.0	0.40
1.0	0.30

ويجدر الانتباه إلى أن تغير $\sqrt{Pq}$ طفيف جدا من أجل قيم P القريبة من 0.5 . لذلك فإنه بتبديل التقدير النقطي $\hat{P} = \frac{X}{n}$ بدلا من p في داخل الجذر نجد أن :

$$P\left(\hat{P} - Z_{\alpha/2}\sqrt{\frac{\hat{P}\,\hat{q}}{n}} < P < \hat{P} + Z_{\alpha/2}\sqrt{\frac{\hat{q}\,\hat{q}}{n}}\right) \simeq 1 - \alpha$$

وهكذا نجد أننا سحبنا عينة ذات حجم n من مجتمع حداني فيه $\hat{P}$ مجهول وحسبنا $\hat{p} = \frac{X}{n}$ من خلال هذه العينة فعندئذ يعتبر المجال :

$$\hat{P} - Z_{\alpha/2}\sqrt{\frac{\hat{P}\,\hat{q}}{n}} < P < \hat{P} + Z_{\alpha/2}\sqrt{\frac{\hat{q}\,\hat{q}}{n}}$$

ممثلا لـ $100\%\ (1 - \alpha)$ مجال ثقة للوسيط P .

ملاحظة

إن المجال السابق هو $(1 - \alpha)100\%$ مجال ثقة لـ p محسوباً من عينة عشوائية حجمها

$n \geq 30$ مختارة من مجتمع حدانى .

مثال (٦,١٠)

أعطيت نتائج استطلاع للرأى حول انتخاب السيد X فى منطقة معينة من بين مائة ناخب ، أن $x = 59$ ناخبا يؤيدون ترشيح المرشح المذكور . لنبحث عن تقدير نسبة الناخبين فى المنطقة الذين يفضلون هذا الناخب ولنضع حدودا لخطأ هذا التقدير .

نلاحظ أن التقدير النقطى لـ P هو $\hat{P} = \frac{59}{100} = 0.59$. وحدود الخطأ المرتكب فى هذا التقدير هى حوالى :

$$2\sqrt{\frac{\hat{p}\,\hat{q}}{n}} = 2\sqrt{\frac{(0.59)\,(0.4)}{100}} = 0.096$$

كما أن 95% مجال ثقة لـ p هو المجال :

$$\left(\hat{p} - Z_{\alpha/2}\sqrt{\frac{\hat{p}\,\hat{q}}{n}}\ ;\ \hat{p} + Z_{\alpha/2}\sqrt{\frac{\hat{p}\,\hat{q}}{n}}\right)$$

وبالتعويض نجد أن :

$$0.59 - (1.96)\,(0.096)\ ,\ 0.59 + (1,96)\,(0.096)$$

أى المجال :

$$0.4018 < p < 0.7781$$

مثال (٦,١١)

لدى دراسة النسبة الفعلية لعدد التلفزيونات الملونة الموجودة فى مدينة جدة بالمملكة العربية السعودية اخترنا عينة عشوائية مؤلفة من 450 عائلة تقتنى أجهزة تلفزيونية ، فوجدنا أن عدد الأجهزة الملونة بينها هو $x = 300$. لنفتش عن 90% مجال ثقة لنسبة التلفزيونات الملونة المستخدمة فى هذه العملية .

إن التقدير النقطى للنسبة p هو التقدير $\hat{p} = \frac{300}{450} = 0.666$ باستخدام الجدول IV

نجد أن :

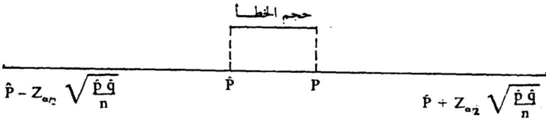
$Z_{0.05} = 1.645$ ، ومن المعلوم أن 90% مجال ثقة للنسبة p هو المجال :

$$\hat{p} - Z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}\hat{q}}{n}} < p < \hat{p} + Z_{\alpha/2} \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

وبالتعويض نجد أن :

$$0.63 < p < 0.70$$

نلاحظ أنه إذا كانت القيمة p هى مركز الـ $(1-\alpha)100\%$ مجال الثقة ، عندئذ يقدر $\hat{p}$ النسبة p دون حدوث أى خطأ . وفى أغلب الأحوال لا تتساوى $\hat{p}$ مع p ، وهذا يعنى أن هناك خطأ مرتكبا فى التقدير ، وحجم هذا الخطأ هو الفرق بين القيمتين p و $\hat{p}$. ويمكننا بثقة قدرها $(1-\alpha)100\%$ من أن نقول بأن هذا الخطأ أقل من $Z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$. ويمكن بسهولة أن نرى ذلك فى المخطط التالى :



نظرية (٦,٣)

إذا كان p تقديرا للنسبة p فى مجتمع حدانى . فيمكن أن نثق بنسبة قدرها $(1-\alpha)100\%$ من أن الخطأ المرتكب فى هذا التقدير سيكون أقل من $Z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$. ففى المثال (٦,١١) كانت نسبة ثقتنا 90% من أن $p = 0.666$ تختلف عن النسبة الحقيقية p بمقدار أقل من 0.035 (الممثلة لنصف طول مجال الثقة) .

لنحدد الآن الحجم الضرورى للعينة المختارة ليكون p تقديرا للنسبة p بخطأ لا يتجاوز العدد $e = Z_{\alpha/2}\sqrt{\hat{p}\hat{q}/n}$ مثلا .

نظرية (٦,٤)

إذا كان $\hat{p}$ تقديرا لـ p فعندئذ يمكن أن نثق بنسبة $(1-\alpha)100\%$ من أن الخطأ المرتكب فى

عملية التقدير هذه لن يتجاوز العدد e إذا كان حجم العينة :

$$n = Z^2_{\alpha/2} \frac{\hat{p}\,\hat{q}}{e^2}$$

مثال (٦,١٢)

ما هو حجم العينة المطلوبة للتأكد من أن الخطأ المرتكب فى تقدير P فى المسألة (٦,١١) لن يبلغ 0.03 بثقة قدرها 90% ؟

من الفرض لدينا $\hat{p} = 0.666$ ، واعتمادا على النظرية (٦,٤) نجد أن :

$$n = \frac{Z^2_{\alpha/2}\hat{p} \cdot \hat{q}}{e^2} = \frac{(1.645)^2 (0.666)(0.334)}{(0.03)^2} = 668.8$$

أى أن $n = 669$ جهاز تلفزيونى . وهذا يعنى أننا إذا أسسنا تقديرنا للنسبة P من خلال عينة حجمها $n = 669$ ، فإنه يمكننا أن نثق بـ 90% من أن تقديرنا $\hat{P}$ لن يختلف عن القيمة الحقيقية P بأكثر من 0.03 .

(٦,٦) تقدير الفرق بين نسبتى مجتمعين حدانيين

Estimating the difference between two proportions

نفرض أننا أمام مجتمعين حدانيين بالنسبة P_1 , P_2 . لإيجاد تقدير لفرق هاتين النسبتين $P_1 - P_2$ مثلاً نختار عينة من المجتمع الأول حجمها n_1 ، وعينة ثانية من المجتمع الثانى حجمها n_2 ، وبصورة تكون معها العينتان مستقلتين ، ثم نحسب من خلال هاتين العينتين نسبة النجاح فى كل عينة ، أى نقوم بحساب $\hat{P}_1$, P_2 ، فيكون الفرق $\hat{P}_1 - \hat{P}_2$ تقديراً نقطياً وغير متحيز للفرق $P_1 - P_2$ ذلك لأن :

$$E(\hat{P}_1 - \hat{P}_2) = P_1 - P_2$$

أما انحرافه المعيارى فيساوى :

$$\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{P_1 q_1}{n_1} + \frac{P_2 q_2}{n_2}}$$

ومن المعلوم أن لكل من $\hat{P}_1 , \hat{P}_2$ توزيعا قريبا من التوزيع الطبيعى بالوسط $P_1 - P_2$ على الترتيب والتباين $\frac{P_2 q_2}{n_2}$ و $\frac{P_1 q_1}{n_1}$. على الترتيب أيضا . وحيث أن العينتين المسحوبتين مستقلتان ، لذلك فإن المتغيرين $\hat{P}_1 , \hat{P}_2$ مستقلان أيضا .

وحسب النظرية (٥,١١) نستنتج أن للمتغير $\hat{P}_1 , \hat{P}_2$ توزيعا قريبا من التوزيع الطبيعى بالوسط $P_1 - P_2$ والتباين $\frac{P_2q_2}{n_2} + \frac{P_1q_1}{n_1}$ لذلك فإن باستطاعتنا التأكد من أن :

$$P \left[- Z_{\alpha/2} < \frac{(\hat{P}_1 - \hat{P}_2) - (P_1 - P_2)}{\sqrt{\frac{p_1 - q_1}{n_1} + \frac{p_2 - q_2}{n_2}}} < Z_{\alpha/2} \right] = 1 - \alpha$$

حيث تمثل $Z_{\alpha/2}$ قيمة المتغير العشوائى الطبيعى المعيارى والتى تحدد على يمينها وتحت منحنى الكثافة مساحة قدرها $\alpha/2$.

وبإجراءات بعض التغيرات فى أطراف المتباينة السابقة نجد أن :

$$p \left(\hat{P}_1 - \hat{P}_2 - Z_{\alpha/2} \sqrt{\frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}} < P_1 - P_2 < P_1 - P_2 + Z_{\frac{\alpha}{2}} \sqrt{\frac{P_1q_1}{n_1} + \frac{P_2q_2}{n_2}} \right) = 1 - \alpha$$

فإذا كان حجما العينتين n_1 , n_2 كبيران . فإن بامكاننا تبديل P_1 , P_2 تحت إشارة الجذر بالتقديرين : $\hat{P}_1 = \frac{x_1}{n_2}$ $\hat{P}_2 = \frac{x_2}{n_2}$ على الترتيب لذلك فإن :

$$P \left(\hat{P}_1 - \hat{P}_2 - Z_{\alpha/2} \sqrt{\frac{\hat{P}_1\hat{q}_1}{n_1} + \frac{\hat{P}_2\hat{q}_2}{n_2}} < P_1 - P_2 + < \hat{P}_1 - \hat{P}_2 \; Z_{\alpha/2} \sqrt{\frac{\hat{P}_1\hat{q}_1}{n_1} + \frac{\hat{P}_2\hat{q}_2}{n_2}} \right)$$

وهكذا نجد أن $(1 - \alpha)100\%$ مجال ثقة للفرق $P_1 - P_2$ هو المجال :

$$\left(\hat{P}_1 - \hat{P}_2 - Z_{\alpha/2} \sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}} , \hat{P}_1 - \hat{P}_2 + Z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}} \right)$$

مثال (٦,١٣)

لمقارنة فاعلية نوعين A , B من المحاليل المضادة للبعوض ، استخدمت غرفتان من نفس الحجم تحتوى كلا منهما على 500 بعوضة ، وعولجت إحداهما بكمية معينة من المحلول A ، أما الثانية فقد عولجت بنفس الكمية السابقة بالمحلول B . وقد وجد أن المحلول A قد أهلك 420 بعوضة ، فى حين أهلك المحلول B 390 بعوضة ، لنفتش عن تقدير الفرق بين قدرتى المحلولين على إبادة البعوض عند استخدامهما فى نفس الشروط المحيطية .

نلاحظ أن تعرض كل بعوضة لهذا المحلول هو تكرار مستقل من مجتمع حدانى . فإذا فرضنا أن احتمال النجاح (أى هلاك بعوضة) هو P_1 بالنسبة للمحلول A و P_2 بالنسبة للمحلول B فعندئذ يكون علينا إيجاد تقدير للفرق $P_1 - P_2$. باستخدام معطيات الحل نجد أن :

$$n_1 = 500 , X_1 = 420 , \hat{P}_1 = \frac{X_1}{n_1} = 0.84$$

$$n_2 = 500 , X_2 = 390 , \hat{P}_2 = \frac{X_2}{n_2} = 0.78$$

وكتقدير للفرق $P_1 - P_2$ نأخذ التقدير النقطى $\hat{P}_1 - \hat{P}_2 = 0.06$ لنفتش عن 0.95 مجال ثقة لهذا التقدير . من الواضح أن 0.95 مجال الثقة هو المجال :

$$\hat{P}_1 - \hat{P}_2 - Z_{\alpha/2} . \sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}} , \hat{P}_1 - \hat{P}_2 + Z_{\alpha/2} . \sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$$

وبالتعويض عن قيمة $Z_{\alpha/2} = Z_{0.025} = 1.96$ من الجدول IV نجد أن 0.95 مجال الثقة لفرق التقديرين هو المجال :

$$(0.06 - 1.96 \sqrt{\frac{(0.84)(0.16)}{500} + \frac{(0.78)(0.22)}{500}} ,$$

$$(0.06 \times 1.96 \sqrt{\frac{(0.84)(0.16)}{500} + \frac{(0.78)(0.22)}{500}})$$

أما حدود خطأ هذا التقدير فهى :

$$(0.06 - 0.048 , 0.06 + 0.048) = (0.012 , 0108)$$

وثقتنا بمثل هذا التقدير ناتجة عن معرفتنا بأنه إذا أعدنا نفس التجربة مرارا وتكراراً ، وحسبنا فى كل مرة تقديراً مجالياً للفرق $P_1 - P_2$ ، فإن 95% تقريبا من هذه المجالات ستحوى القيمة الحقيقية للفرق $P_1 - P_2$.

(٦,٧) تقدير التباين Estimating the variance

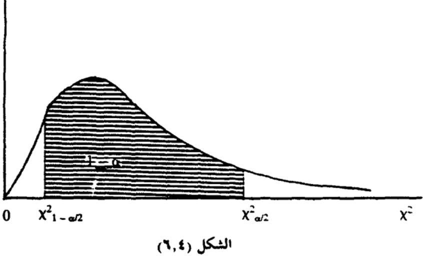
كتقدير نقطى لتباين مجتمع إحصائى σ^2 نأخذ عادة تباين العينة S^2 والذى يمثل كما أوضحنا تقديراً غير متحيز لتباين المجتمع المجهول σ^2 $E(S^2 = \sigma^2)$ من المعلوم أن توزيع الإحصاء :

$$X^2 = \frac{(n - 1)\, S^2}{\sigma^2}$$

هو التوزيع $\chi^2_{(n-1)}$ (كاى – مربع) بـ $(n - 1)$ درجة من الحرية النظرية (٥,١٦) ، وذلك بفرض أن العينة التى حسبنا منها الإحصاء السابق قد أخذت من مجتمع طبيعى تباينه σ^2 . لذلك يمكن أن نكتب :

$$P(\chi^2_{1-\alpha/2} < X^2 < \chi^2_{\alpha/2}) = 1 - \alpha$$

حيث تمثل $\chi^2_{\alpha/2}$ قيمة كاى – مربع بـ $(n - 1)$ درجة حرية ، والتى تحدد على يمينها وتحت منحنى كثافة χ^2 مساحة قدرها $\alpha/2$ انظر الشكل (٦,٤)



الشكل (٦,٤)

كما نلاحظ على الشكل (٦،٤) أن $\chi^2_{1-\alpha/2}$ تمثل قيمة χ^2 بـ $n-1$ درجة حرية ، والتى تحدد على يمينها مساحة قدرها $1-\alpha/2$.

وبتبديل قيمة الإحصاء $X^2 = \frac{(n-1)}{\sigma^2} S^2$ فإننا نجد أن :

$$P\left[\chi^2_{1-\alpha/2} < \frac{(n-1)\,S^2}{\sigma^2} < \chi^2_{\alpha/2}\right] = 1-\alpha$$

وبتقسيم كافة عناصر المتباينة على $(n-1)S^2$ ثم بأخذ مقلوب هذه المتباينة المزدوجة نجد أن :

$$P\left[\frac{(n-1)S^2}{\chi^2_{\alpha/2}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{1-\alpha/2}}\right] = 1-\alpha$$

وأخيراً فإن $(1-\alpha)100\%$ مجال ثقة للتباين المجهول σ^2 فى مجتمع إحصائى هو المجال :

$$\left(\frac{n-1}{\chi^2_{\alpha/2}} S^2 \;,\; \frac{n-1}{\chi^2_{1-\alpha/2}} S^2\right)$$

مثال (٦،١٤)

سحبنا عينة عشوائية حجمها $n = 20$ من مجتمع طبيعى . فأعطت وسط عينة $\bar{x} = 32.8$ وتباين عينة $S = 4.51$. لنحسب 0.95 مجال ثقة للتباين σ^2 .

الحل

نلاحظ أن $1-\alpha = 0.95$. أى أن $\frac{\alpha}{2} = 0.025$. كما نلاحظ أنه من أجل $n-1 = 19$ أن $\chi^2_{0.025} = 32.852$ وأن $\chi^2_{0.975} = 8.907$. كذلك فإن $S = 4.51$.
وهكذا نجد أن 0.95 مجال ثقة للتباين σ^2 هو المجال :

$$\left(\frac{n-1}{\chi^2_{\alpha/2}} S^2 \;,\; \frac{n-1}{\chi^2_{1-\alpha/2}} S^2\right) \quad \left(\frac{19}{32.852} 4.51 \;,\; \frac{19}{8.907} 4.51\right)$$

أى المجال :

$$(2.608 \;,\; 9.620)$$

(٦,٨) تقدير نسبة تباينين Estimating the Ratio of two variances

كتقدير نقطى لنسبة تباينى مجتمعين إحصائيين $\frac{\sigma_1^2}{\sigma_2^2}$ نأخذ عادة النسبة $\frac{S_1^2}{S_2^2}$ أى نسبة تباينى العينتين المسحوبتين من هذين المجتمعين . لذلك فإن الإحصاء $\frac{S_1^2}{S_2^2}$ يشكل تقديراً لـ $\frac{\sigma_1^2}{\sigma_2^2}$. إذا كان المجتمعان المدروسان طبيعيين فعندئذ يمكن إنشاء مجال لتقدير $\frac{\sigma_1^2}{\sigma_2^2}$ وذلك باستخدام الإحصاء .

$$F = \frac{\sigma_2^2 S_1^2}{\sigma_1^2 S_2^2}$$

وبحسب النظرية (٥,٢٠) فإن المتغير العشوائى F له توزيع F بعدد من درجات الحرية $\nu_1 = n_1 - 1$ ، $\nu_2 = n_2 - 1$. لذلك واعتمادا على الشكل (٦,٥) يمكننا أن نكتب :

$$P[\ f_{1-\alpha/2}(\nu_1\ ,\ \nu_2) < F < f_{\alpha/2}(\nu_1\ ,\ \nu_2)\ ] = 1-\alpha$$

حيث تمثل $f_{\alpha/2}(\nu_1 , \nu_2)$ قيمة F بـ ν_1 , ν_2 درجات حرية ، والتى تحدد على يمينها مساحة قدرها $\frac{\alpha}{2}$ ، انظر الشكل (٦,٥) .

وبتعويض قيمة F فى العلاقة السابقة نجد أن :

$$P\left[\ f_{1-\alpha/2}(\nu_1\ ,\ \nu_2) < \frac{\sigma_2^2\ S_1^2}{\sigma_1^2\ S_2^2} < f_{\alpha/2}(\nu_1\ ,\ \nu_2)\ \right] = 1-\alpha$$

الشكل (٦,٥)

أى أن :

$$P\left[\frac{S_1^2}{S_2^2}\,\frac{1}{f_{\alpha/2}(v_1\,,\,v_2)}<\frac{\sigma_1^2}{\sigma_2^2}<\frac{S_1^2}{S_2^2}\,\frac{1}{f_{1-\alpha/2}(v_1\,,\,v_2)}\right]=1-\alpha$$

وحيث إن $f_{1-\alpha/2}(\nu_1\,,\,\nu_2)=\dfrac{1}{f_{\alpha/2}(\nu_2\,,\,\nu_1)}$ لذلك فإننا نجد :

$$P\left[\frac{S_1^2}{S_2^2}\,\frac{1}{f_{\alpha/2}(\nu_1\,,\,\nu_2)}<\frac{\sigma_1^2}{\sigma_2^2}<\frac{S_1^2}{S_2^2}\,f_{\alpha/2}(\nu_2\,,\,\nu_1)\right]=1-\alpha$$

وهكذا نستنتج بأنه $(1-\alpha)100\%$ مجال ثقة للنسبة $\dfrac{\sigma_1^2}{\sigma_2^2}$ هو المجال :

$$\frac{S_1^2}{S_2^2}\,\frac{1}{f_{\alpha/2}(\nu_1\,,\,\nu_2)}<\frac{\sigma_1^2}{\sigma_2^2}<\frac{S_1^2}{S_2^2}\,f_{\alpha/2}(\nu_2\,,\,\nu_1)$$

وذلك بفرض أننا سحبنا عينتين مستقلتين ذواتا حجمين $n_1\,,\,n_2$ من مجتمعين طبيعيين لهما تباينان σ_2^2 و σ_1^2 .

مثال (٦,١٥)

أقامت جامعة الملك عبد العزيز بجدة مسابقة بين طلاب كليتى العلوم والهندسة وتقدم 25 طالبا من كلية العلوم إلى المسابقة ، ومن كلية الهندسة 16 طالبا وفى اختبار الرياضيات قدم طلاب كلية العلوم هذا الاختبار بوسط قدره 82 وانحراف معيارى 8 . بينما قدم طلاب كلية الهندسة الاختبار بوسط 78 وانحراف معيارى قدره 7 . لننشىء 98% مجال ثقة للنسبة $\dfrac{\sigma_1^2}{\sigma_2^2}$ ثم $\dfrac{\sigma_1}{\sigma_2}$ ، وذلك بفرض أن σ_1^2 و σ_2^2 يمثلان تبايني مجتمعي درجات كل من طلاب كليتى العلوم والهندسة على الترتيب والمتقدمين إلى مثل هذا الاختبار .

نلاحظ أن :

$$n_1=25\,,\,n_2=16\quad,\quad S_1=8\,,\,S_2=7$$

ومن أجل 98% مجال ثقة نلاحظ أيضا أن $\alpha = 0.02$. وباستخدام الجدول VII نجد أن :

$$f_{0.01}(24, 15) = 3.29 \quad , \quad f_{0.01}(15, 24) = 2.89$$

وبالتبديل فى العلاقة :

$$\frac{S_1^2}{S_2^2} \frac{1}{f_{\alpha/2}(\nu_1 , \nu_2)} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{S_1^2}{S_2^2} f_{\alpha/2}(\nu_2 , \nu_1)$$

فإننا نجد أن :

$$\frac{(64)}{(49)} \frac{1}{(3.29)} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{(64)}{(49)} (2.89)$$

وأخيراً فإن 0.98 مجال ثقة لنسبة التباينين هو المجال :

$$0.397 < \frac{\sigma_1^2}{\sigma_2^2} < 3.775$$

وبأخذ جذرى أطراف المتباينة السابقة نحصل على 0.98 مجال ثقة لـ $\frac{\sigma^1}{\sigma^2}$ هو المجال :

$$0.630 < \frac{\sigma_1}{\sigma_2} < 1.943$$

تمارين محلولة

تمرين (١)

بفرض أن وسط العينة والانحراف المعيارى المحسوبين من خلال عينة حجمها $n = 36$ مسحوبة من مجتمع من الرجال البالغين هما على الترتيب 2.6 , 0.3 . أوجد 95% ثم 99% مجالات ثقة للوسط μ المجهول فى هذا المجتمع .

الحل

نلاحظ أن $\overline{x} = 2.6$ هو تقدير نقطى للوسط المجهول μ فى هذا المجتمع ، وبما أن حجم العينة المسحوبة $n = 36$ ، لذلك يمكن أن نقرب التباين σ^2 إلى S^2 المحسوبة من خلال العينه واعتبار أن $\sigma = S = 0.3$.

أما بالنسبة للنقطة $Z_{\alpha/2}$ فهى النقطة من محور السينات التى تحصر لنا على الترتيب مساحة قدرها 0.025 على يمينها أو 0.975 على يسارها من أجل 95% ثقة ، و 0.005 على يمينها أو 0.995 على يسارها من أجل 99% . وهكذا نجد أن 95% مجال الثقة للوسط μ هو المجال :

$$2.6 - (1.96) \left(\frac{0.3}{\sqrt{36}} \right) < \mu < 2.6 + (1.96) \left(\frac{0.3}{\sqrt{36}} \right)$$

أى أن :

$$2.5 < \mu < 2.71$$

كما نلاحظ أن 99% مجال ثقة لنفس الوسط هو المجال :

$$2.6 - (2.575) \left(\frac{0.3}{\sqrt{36}} \right) < \mu < 2.6 + (2.575) \left(\frac{0.3}{\sqrt{36}} \right)$$

أى أن :

$$2.47 < \mu < 2.73$$

وهكذا نستنتج أن مجال الثقة الأطول يعطى تقديراً بدرجة أعلى من الدقة .

تمرين (٢)

بفرض أن $\tilde{S}^2 = \sum_{i=1}^{n} (X_i - \bar{X})^2 / n$.

برهن أن $E(\acute{S}^2) = \frac{n-1}{n} \sigma^2$ ·

وأن S^2 هو تقدير متحيز لـ σ^2 .

الحل

من الواضح أن :

$$\frac{n\acute{S}^2}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2 / (n-1)$$

$$\frac{n\acute{S}^2}{n-1} = S^2$$

ثم أن :

$$E\left(\frac{n\acute{S}^2}{n-1}\right) = E(S^2)$$

وبما أن S^2 هو تقدير غير متحيز لـ σ^2 مثال (٦,٢) لذلك فإن :

$$E\left(\frac{n\acute{S}^2}{n-1}\right) = \sigma^2$$

وهكذا نجد أن :

$$E\,\acute{S}^2 = \left(\frac{n-1}{n}\right) \sigma^2$$

وحيث أن : $E(\hat{S}_2) = \sigma^2$

لذلك لا يمثل S^2 تقديرا غير متحيز لـ σ^2 فهو يمثل تقديراً متحيزاً لهذا التباين .

تمرين (٣)

المطلوب تقدير وسط الإنتاج اليومى فى منشأة للصناعات الكيميائية ، وذلك بفرض أن وسط عينة من الإنتاج اليومى وانحرافها المعيارى لفترة $n = 50$ يوماً هما على التالى :

$$\bar{X} = 871$$
$$S = 21$$

الحل

من المعلوم أن التقدير الأفضل هو $\bar{X} = 871$ طن يومياً ، وحدود الخطأ فى هذا التقدير هى :

$$\pm 2\sigma_{\bar{x}} = \pm 2 \frac{\sigma}{\sqrt{n}} = 2 \frac{\sigma}{\sqrt{50}}$$

ومع أن σ مجهول ، إلا أنه يمكن اعتبار S كقيمة تقريبية له أى اعتبار S تقديراً لـ σ . وهكذا يكون حدود الخطأ فى هذا التقدير بصورة تقريبية مساويا لـ :

$$\frac{2\,S}{\sqrt{50}} = \frac{2\,(21)}{\sqrt{50}} = 5.94$$

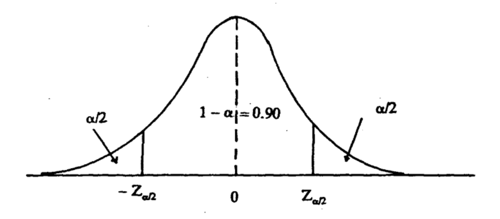
وسنشعر إلى حد ما بالثقة من أن التقدير 871 هو فى حدود 5.94 طن من القيمة الحقيقية لوسط الإنتاج .

تمرين (٤)

احسب 90% مجال ثقة لوسط الإنتاج اليومى فى التمرين (٦,٣) .

الحل

نلاحظ أن :



إن $Z_{\alpha/2}$ الموافقة لأمثال ثقة 90% هى القيمة $Z_{\alpha/2} = 1.645$ ، ومنه فمجال الثقة المطلوب هو المجال :

$$\overline{x} \pm 1.645 \frac{\sigma}{\sqrt{n}}$$

فإذا وضعنا S بدلا عن σ فإننا نحصل بصورة تقريبية على المجال :

$$871 \pm (1.645) \frac{21}{\sqrt{50}}$$

أو :

$$871 \pm 4.89$$

وهكذا نقدر أن وسط الإنتاج اليومى μ يقع ضمن المجال من 866.11 إلى 875.89 طنا . إن أمثال الثقة السابقة تعنى أن 90% من مجالات الثقة الناتجة عن سحب عينات بصورة متكررة ستحوى μ .

تمرين (٥)

احسب 95% مجال ثقة لوسط سعة أو عينة تحتوى على حمض الكبريت إذا علمت أن سبعة منها متشابهة وتتسع على التتالى :

9.8 , 10.2 , 10.4 , 9.8 , 10 , 10.2 , 9.6

الحل

نلاحظ أن وسط العينة والانحراف المعيارى لهذه العينة من الأوعية هما على التتالى :

$$\bar{x} = 10.0 \quad , \quad s = 0.283$$

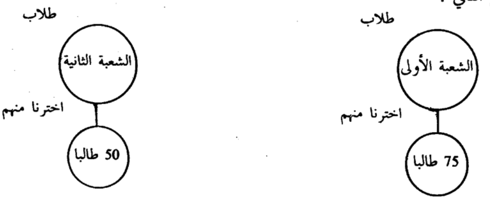
باستخدام الجدول V نجد أن $t_{0.025} = 2.447$ من أجل $\nu = (n-1) = 7-1 = 6$ درجة من الحرية . لذلك فإن 95% مجال ثقة للوسط μ (الممثل لوسط سعة الأوعية المستخدمة) هو المجال :

$$10.0 - (2.447)\left(\frac{0.283}{\sqrt{7}}\right) < \mu < 10.0 + (2.447)\left(\frac{0.283}{\sqrt{7}}\right)$$

$$9.74 < \mu < 10.26$$

تمرين (٦)

فى الاختبار النهائى لمادة الاحتمالات تقدمت مجموعتان تحويان 75 طالبا من الشعبة الأولى و 50 طالبا من الشعبة الثانية . وقد قدم طلاب المجموعة الأولى هذا الامتحان بوسط 82 درجة وانحراف معيارى قدره 8 درجة ، أما طلاب المجموعة الثانية فقد قدموا امتحانهم هذا بوسط قدره 76 درجة وانحراف معيارى 6 . والمطلوب إيجاد 96% مجال ثقة للفرق $\mu_1 - \mu_2$ حيث يمثل μ_1 وسط الدرجات التى أحرزها كل طلاب الشعبة الأولى ، أما μ_2 فتمثل وسط الدرجات التى أحرزها كل طلاب الشعبة الثانية . لاحظ الشكل التالي .



الحل

من شروط المسألة نجد أن :

$$\bar{x}_1 = 82 \qquad s_1 = 8$$
$$\bar{x}_2 = 76 \qquad s_2 = 6$$

نلاحظ أن التقدير النقطي لفرق الوسطين $\mu_1 - \mu_2$ هو : $\bar{X}_1 - \bar{X}_2 = 82 - 76 = 6$
كما نلاحظ أن حجم العينة الأولى $n_1 = 75$ ، وحجم العينة الثانية $n_2 = 50$. لذلك يمكن استبدال σ_2 و σ_1 على الترتيب بـ S_2 و S_1 وحيث إن $1 - \alpha = 0.96$ لذلك $\alpha = 0.04$. ومن الجدول IV نجد أن $Z_{\alpha/2} = Z_{0.02} = 2.054$. وهكذا نجد أن 96% مجال ثقة لفرق الوسطين هو المجال :

$$\left((\bar{x}_1 - \bar{x}_2) - Z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \, , \, (\bar{x}_1 - \bar{x}_2) + Z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right)$$

أى :

$$\left(6 - 2.054 \sqrt{\frac{64}{75} + \frac{36}{50}} \, , \, 6 + 2.054 \sqrt{\frac{64}{75} + \frac{36}{50}} \right)$$

$$3.42 < \mu_1 - \mu_2 < 8.58$$

تمرين (٧)

عينة عشوائية مؤلفة من قياسات قطر جسم كروى كانت قد سجلت على النحو التالى :

$$3.37 \; , \; 6.36 \; , \; 6.37 \; , \; 6.35 \; , \; 6.34 \text{ mm}$$

أوجد تقديراً غير متحيز للوسط المجهول ثم للتباين المجهول أيضا .

الحل

نلاحظ أن وسط العينة يشكل تقديراً غير متحيز للوسط المجهول ، ولذلك فإن :

$$\bar{x} = \frac{\sum_{i=1}^{5} x_i}{5} = \frac{6.37 + 6.36 + 6.37 + 6.35 + 6.34}{5} = 6.35 \text{ mm}$$

ونعلم أن تباين العينة S^2 يشكل تقديراً غير متحيز للتباين σ^2 كما أن :

$$S^2 = \frac{\sum_{i=1}^{5} (x_i - \bar{x})^2}{5 - 1}$$

$$\frac{(6.37 - 6.35)^2 + (6.36 - 6.35)^2 + (6.37 - 6.35)^2 + (6.34 - 6.35)^2 + (6.35 - 6.35)^2}{4} = S^2$$

$$S^2 = 0.00025 \text{ (mm)}^2$$

تمرين (٨)

لقياس زمن رد الفعل ، قدر أحد العلماء الانحراف المعيارى بـ 0.05 ثانية . ما هو حجم العينة من القياسات بحيث تكون بـ 95% . واثقين من أن خطأ التقدير لن يتجاوز 0.01 ثانية ؟

الحل

من الفرض لدينا $1 - \alpha = 0.95$

ولذلك فإن : $Z_{\alpha/2} = Z_{0.025} = 1.96$

كما أن : $\sigma = S = 0.05$ ثانية

ونعلم أن الخطأ e يعطى بالعلاقة :

$$e \leq \frac{\sigma}{\sqrt{n}} Z_{\alpha/2}$$

ومن هذا نجد أن :

$$n \leq \left(\frac{\sigma}{e} Z_{\alpha/2} \right)^2$$

$$n \leq \left(1.96 \, \frac{0.05}{0.01} \right)^2$$

أى أن :

$$n \leq 96.04$$

$$n = 96$$

تمرين (٩)

تم استطلاع الرأى العام فى جامعة الملك عبد العزيز حول انتخاب المرشح محمد بن المبارك لشغل منصب أمين عام العلاقات العامة . فاختيرت عينة من المسئولين ممن يحق لهم الانتخاب حجمها 100 شخص وتم استجوابهم حول ترشيحهم للسيد المذكور فدلت النتائج على أن 57% منهم يؤيدون ترشيح المرشح المذكور . أوجد 99% مجال ثقة للنسبة بين جميع الناخبين المؤيدين للسيد محمد بن المبارك .

الحل

نفرض أن P يمثل نسبة المؤيدين للمرشح . ومن الفرض لدينا :

$$1 - \alpha = 0.99$$

لذلك فإن :

$$\alpha = 0.01 \rightarrow Z_{\alpha/2} = Z_{0.005}$$

نلاحظ من الجدول IV أن النقطة $Z_{0.005}$ التى تحصر على يسارها تحت منحنى الكثافة الطبيعى المعيارى مساحة قدرها 0.995 هى النقطة :

$$Z_{\alpha/2} = 2.58$$

ثم إن :

$$\hat{P} = 0.55$$

$$\hat{q} \geqslant 1 - \hat{P} = 0.45$$

$$n = 100$$

ونعلم أن $(1 - \alpha)100\%$ مجال الثقة للنسبة P هو المجال :

$$\left(\hat{P} - Z_{a/2} \sqrt{\frac{\hat{P}\,\hat{q}}{n}} \;,\; \hat{P} + Z_{a/2} \sqrt{\frac{\hat{P}\,\hat{q}}{n}} \right)$$

وبالتعويض نجد أنه :

$$(0.55 - 2.58) \sqrt{\frac{(0.55)\,(0.45)}{100}} \quad , 0.55 + 2.58 \sqrt{\frac{(0.55)\,(0.45)}{100}} \;)$$

وأخيراً فإن مجال الثقة هو :

$$(0.4216 \quad , \quad 0.6783)$$

تمرين (١٠)

ينتج مصنعان A و B نوعين متشابهين من اللمبات الكهربائية . أخذت عينة من إنتاج المصنع A حجمها $n_1 = 150$ لمبة وجربت ، فكان وسط عمر كل لمبة من هذه اللمبات 1400 ساعة عمل وانحرافها المعيارى 120 ساعة ، كما أخذت عينة أخرى إنتاج المصنع B حجمها $n_2 = 200$ لمبة وجربت وسجلت نتائج التجربة . فإذا علمت أن وسط عمر كل لمبة كان 1200 ساعة وانحرافها المعيارى 80 ساعة ، فأوجد 95% مجال ثقة للفرق بين وسط العمر للمبات المصنعين A و B .

الحل

نعلم أن : $1 - \alpha = 0.95$

وهذا يعنى : $Z_{\alpha/2} = 1.96$

لدينا أيضا : $\bar{x}_1 = 1400 \quad , \quad \sigma_1^2 = (120)^2$

$\bar{x}_2 = 1200 \quad , \quad \sigma_2^2 = (80)^2$

فإذا فرضنا أن وسط عمر أية لمبة من إنتاج المصنع A هو μ_1 وأن μ_2 هو وسط عمر أى لمبة من إنتاج المصنع B . فمن المعلوم أن 95% مجال ثقة لفرق الوسطين $\mu_1 - \mu_2$ هو المجال :

$$\left(\bar{x}_1 - \bar{x}_2 - \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \cdot Z_{\alpha/2} \;,\; \bar{x}_1 - \bar{x}_2 + \sqrt{\frac{\sigma_1^2}{n_1} = \frac{\sigma_2^2}{n_2}} \cdot Z_{\alpha/2} \right)$$

وبالتعويض نجد أن :

$$1200 - 1.96\sqrt{\frac{(120)^2}{150} + \frac{(80)^2}{200}} \quad , 1400 - 1200 + 1.96\sqrt{\frac{(120)^2}{150} + \frac{(80)^2}{200}}\ \Big)$$

$$\Big(1400 -$$

أى أن :

تمرين (١١)

عشرة رزم من حبوب الأعشاب أوزانها بالديسغرام :

46.9 , 45.2 , 46.4 , 46.0 , 46.1 , 45.8 , 47.0 , 46.1 , 45.9 , 45.8

أوجد 0.95 مجال ثقة لتباين كل رزم الأعشاب السابقة .

الحل

$$S^2 = \frac{n\sum_{i=1}^{n} X_i^2 - (\sum_{i=1}^{n} X_i)}{n(n-1)} = \frac{10\,(21273.12) - (461.2)^2}{(10)\,(9)} = 0.286$$

وللحصول على 0.95 مجال ثقة نلاحظ أن $0.05 = \alpha$. ومن الجدول IV نجد أن $19.023 = {}^{2}_{0.025}$ وأن $2.70 = {}^{2}_{0.975}$ وذلك من أجل $v = 9$ ، وبالتعويض فى العلاقة :

$$\frac{(n-1)S^2}{\chi^2_{\alpha/2}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{1-\ /2}}$$

نجد أن :

$$\frac{(9)\,(0.286)}{19.023} < \sigma^2 < \frac{(9)\,(0.286)}{2.70}$$

$$0.135 < \sigma^2 < 0.953$$

تمارين عامة

(١) ينتج مصنعا للمصابيح الكهربائية نوعا من النيون ، وبفرض أن μ يمثل وسط عمر أى نيون من إنتاج المصنع ، وبقصد البحث عن تقدير لوسط العمر μ اخترنا بصورة عشوائية عينة من مئتى نيون فأعطت وسط عمر قدره 1500 ساعة وانحراف معيارى قدره 143 ساعة ، والمطلوب البحث عن تقدير غير متحيز لـ μ ، ووضع حدود لخطأ هذا التقدير .

(٢) اخترنا عينة عشوائية مؤلفة من 300 ترانزيستور وجربناها ، فوجدنا أنها تحتوى على 30 ترانزيستورا عاطلا . أنشىء 0.90 مجال ثقة لنسبة الترانزيستورات العاطلة .

(٣) إذا علمت أن وسط عينة مؤلفة من 50 قياسا هو 36.5 ، وأن انحرافها المعيارى هو 2.1 . فالمطلوب إنشاء 0.98 مجال ثقة لوسط المجتمع الذى أخذت منه هذه العينة ، ضع حدودا لخطأ هذا التقدير .

(٤) ترغب إدارة مستشفى جامع الملك عبد العزيز بجدة فى تقدير عدد الأيام التى يحتاجها علاج مرضى الجرب ، والذين يتراوح سنهم بين 15 و 50 عاما . ولهذا الغرض اختيرت عينة عشوائية من المرضى مؤلفة من 500 مريضاً . فوجد أن وسط عدد الأيام فى هذه العينة هو 5.3 يوما ، وأن انحرافها المعيارى هو 1.25 يوما . والمطلوب تقدير وسط الإقامة فى المستشفى المذكور لجميع مرضى الجرب الذين اختيرت منهم هذه العينة .

(٥) استخدم عمال الحفر للتنقيب عن البترول نوعين من الحفارات من إنتاج مصنعين مختلفين A و B ، وقد وجدوا أن وسط اختراق الحفارة الأولى (من إنتاج المصنع A) هو 10.8 ، وأن انحرافها المعيارى هو 1.2 بوصة . وذلك عندما قاموا بحفر 50 حفرة . كما وجدوا أن وسط اختراق الحفارة من النوع الثانى هو 9.1 وانحرافها المعيارى 1.6 بوصة وذلك عند حفر 80 حفرة صخرية أخرى . والمطلوب تقدير الفرق فى معدل الاختراق الوسطى لهاتين الحفارتين إذا علمت أن بنية التربة ومدة الحفر واحدة بالنسبة للحفارتين .

(٦) تم سحب عينتين مستقلتين حجم كل منهما $n = 64$ من مجتمعين طبيعيين لهما نفس الوسط 6.4 ونفس الانحراف المعيارى 7.2 . ما هو احتمال أن يتجاوز الفرق بين وسطى العينيين العدد 90.6 بالقيمة المطلقة ؟

(٧) اختيرت عينتان عشوائيتان حجماهما $n_2 = 16 , n_1 = 9$ من مجتمعين طبيعيين مستقلين بالوسطين العينيين $\overline{x}_1 = 64 , \overline{x}_2 = 59$ والانحرافين المعياريين $S_1 = 6 , S_2 = 5$ أنشىء 95% مجال ثقة للفرق $\mu_1 - \mu_2$ وذلك بفرض أن $\sigma_1 = \sigma_2$.

(٨) فى مدينة معينة اختيرت عينة من 200 شخص بصورة عشوائية . ووجد أن 114 منهم يؤيدون ترشيح المرشح X . أنشىء 96% مجال ثقة لنسبة المؤيدين لهذا المرشح فى المدينة المذكورة .

(٩) اختير 500 شخص من المدخنين فى مدينة جدة بالمملكة العربية السعودية وتم سؤالهم عن رأيهم حول نوع معين من السجاير فوجد أن 86 شخصا منهم يفضلون هذا النوع . أنشىء 90% مجال ثقة لنسبة المدخنين الذين يفضلون هذا النوع من السجاير .

(١٠) سحبنا عينة عشوائية حجمها $n = 100$ عنصر من جملة بضاعة مصنعة جاهزة للتسويق وفحصناها فوجدنا أن هذه العينة تحتوى على ثمانية عناصر معيبة ، أنشىء 98% مجال ثقة لنسبة العناصر المعيبة .

(١١) تحاول شركة لتأجير السيارات شراء إطارات لسيارتها (التى تعمل بين مكة المكرمة ــ جدة) من أحد النوعين A و B من الإطارات ، ولتقدير الفرق بين النوعين اختارت هذه الشركة اثنى عشر إطارا من كل نوع ووضعتهما تحت الخدمة حتى بليت جميعها . وقد أعطت التجربة السابقة المعلومات التالية حول هذين النوعين ، فالبنسبة للنوع A أوضحت التجربة أن ($S_1 = 5000$ كيلومترا) وأن ($\overline{x}_1 = 36300$ كيلومترا)
أما بالنسبة للنوع B فأوضحت أن ($S_2 = 6100$ كيلومتراً) ، وأن ($\overline{x}_2 = 38100$ كيلومتراً)
والمطلوب إنشاء 95% مجال ثقة للفرق $\mu_2 - \mu_1$ ، وذلك بفرض أن لكل من مجتمعى الاطارات السابقة توزيعاً قريباً من التوزيع الطبيعي .

(١٢) أنشىء 90% مجال ثقة لـ σ_2^2 / σ_1^2 فى المثال السابق .

الفصل السابع

اختبارات الفرضيات

■ الفرضية الاحصائية ■ الأخطاء من النوع الأول I ومن النوع الثانى II ■ الاختبارات وحيدة وثنائية الذيل ■ اختبار وسط وتباين مجتمع احصائى ■ اختيار حجم العينة لاختبار الوسط ■ الاختبارات المتعلقة بالنسب ■ اختبار الفرق بين نسبتين ■ تمارين محلولة.



(٧,١) الفرضية الإحصائية

كثيرا ما نلجأ إلى اتخاذ قرار يخص مجتمعاً مدروساً وذلك بناء على معلومات تقدمها العينة الاحصائية ، مثل هذا القرار يدعى عادة بالقرار الاحصائى . وبغية الحصول على قرار مفيد لابد من وضع فروض أو تخمينات تتعلق أيضا بهذا المجتمع الذى ندرسه ، مثل هذه الفروض قد تكون صحيحة أو غير صحيحة وتدعى بالفروض الإحصائية .

ويتألف أى اختبار إحصائى من أربعة عناصر :

١ ــ فرض العدم
٢ ــ إحصاء الاختبار
٣ ــ منطقة الرفض
٤ ــ الفرضية البديلة

إن تحديد هذه العناصر الأربعة اختباراً معيناً ، وتغيير عنصر أو أكثر يؤدى إلى اختبارات جديدة .

(٧,٢) الأخطاء من النوع الأول I ومن النوع الثاني II

Errors: kind I and II

نرمز عادة لفرض العدم بالرمز Ho ، وهذا الفرض يحدد قيماً افتراضية لوسيط أو أكثر من وسطاء المجتمع . فقد نرغب مثلا فى اختبار الفرض بأن نوعاً معيناً من لقاح لجدرى الماء فعالا خلال عامين بنسبة 25% .

إن قرارنا برفض أو قبول فرض العدم Ho مبنى على المعلومات التى تحويها عينة مسحوبة من المجتمع الإحصائى المدروس . وتستخدم قيم العينة لحساب عدد واحد يأخذ

دور صانع القرار ، ويدعى بإحصاء الاختبار . ونقسم مجموعة كل القيم التى يمكن أن يأخذها إحصاء الاختبار إلى مجموعتين أو منطقتين تدعى إحداهما بمنطقة الرفض ، كما تدعى الثانية بمنطقة القبول . فإذا وقعت القيمة التى يأخذها إحصاء الاختبار فى منطقة الرفض ، فإننا نرفض الفرضية الابتدائية . ونقبلها إذا وقعت فى منطقة القبول .

وتخضع طريقة اتخاذ القرار هذه لنوعين من الخطأ . إذ يمكن أن نرفض فرض العدم ، بينما هو فى الواقع صحيح ، أو يمكن أن نقبل Ho وهو فى الحقيقة غير صحيح ، والصحيح هو فرض آخر بديل . يدعى هذان النوعان من الأخطاء بالخطأ من النوع الأول I ، والخطأ من النوع الثانى II على الترتيب . أى أن رفض فرض صحيح (أو الرفض الخاطىء) هو الخطأ من النوع الأول ، وقبول فرض غير صحيح (أى القبول الخاطىء) هو الخطأ من النوع الثانى . ويوضح الجدول (٧,١) الحالتين الممكنتين لفرض العدم ، ونوعى القرار الممكنين ، بالإضافة إلى نوع الخطأ الذى يمكن ارتكابه .

القرار	فرض العدم	
	صحيح	مخطىء
الرفض	الخطأ من النوع الأول I	قرار صحيح
القبول	قرار صحيح	الخطأ من النوع الثانى II

الجدول (٧,١)

ونقتبس جودة الاختبار الإحصائى لفرض باحتمالى الخطأ من النوع الأول I ، والخطأ من النوع الثانى II ، ونرمز لهما بالرمزين α , β على التوالى . ويمكن التعبير عن الاحتمال α بأنه احتمال أن يقع إحصاء الاختبار فى منطقة الرفض علما بأن Ho صحيح ، تدعى α أيضا بمستوى معنوية الاختبار . ومن الواضح أن زيادة حجم منطقة الرفض سيزيد من قيمة α . وفى نفس الوقت سيؤدى إلى تناقص β . وذلك من أجل قيمة ثابتة لـ n ، أما تخفيض حجم منطقة الرفض فسيؤدى إلى تناقص α وذلك من أجل قيمة ثابتة لـ n ، أما تخفيض حجم منطقة الرفض فسيؤدي إلى تناقص α وزيادة β ، وذلك عند زيادة حجم العينة n . فمن الطبيعى أن تقدم لنا العينة الأكبر حجما ، قدرا أكبر من المعلومات نتخذ على ضوئها قرارنا ، وهذا يؤدى إلى تناقص كل من α و β .

لنفرض على سبيل المثال أن لقاحاً معيناً ضد جدرى الماء فعال بنسبة 20% بعد عامين من استخدامه ، لنفرض أن شركة للأدوية اكتشفت لقاحاً جديداً ضد نفس المرض ، وتريد معرفة فعالية هذا الدواء لفترة زمنية أكبر . لذلك اختارت هذه الشركة بصورة عشوائية 15 شخصاً وتم تلقيحهم بهذا اللقاح الجديد . وتعتبر الشركة أن اللقاح الجديد هو أكثر فعالية إذا صدف أن سبعة أو أكثر من هؤلاء الأشخاص قد مر على تلقيحهم عامين أو أكثر دون التقاطهم لفيروس هذا المرض .

نلاحظ أن العدد 7 كيفى إلى حد ما ، ولكن يبدو معقولاً لأنه يمثل عدداً أكبر من ثلاثة وهو عدد الأشخاص الذين يتوقع أن يحصلوا على مقاومة ضد المرض فيما إذا تم تلقيح 15 شخصا باللقاح القديم . لنختبر الآن الفرض العدم فى أن المتغير الحدانى (وهو احتمال النجاح فى اختبار ما) هو $P = \frac{1}{5}$ ضد الفرض البديل فى أن $P > \frac{1}{5}$ وهذا ما يكتب عادة على النحو التالى :

$$H_0 : P = \frac{1}{5}$$

$$H_0 : P > \frac{1}{5}$$

ويمثل الإحصاء X الذى بنى عليه قرارانا عدد الأشخاص فى عينتنا ، والذين حصلوا باستخدام اللقاح الجديد على مقاومة لمدة عامين أو أكثر . لنجزىء مجموعة القيم الممكنة من صفر إلى 15 إلى مجموعتين : تمثل الأولى مجموعة الأعداد الأقل من 7 ، أما الثانية فتمثل مجموعة الأعداد الأكبر من سبعة أو تساويها ، هذا وتؤلف مجموعة النقاط الأكبر من 6.5 منطقة تدعى بالمنطقة الحرجة ، كما تمثل مجموعة النقاط الأقل من 6.5 منطقة القبول 9 ، والعدد 6.5 الذى يفصل هاتين المنطقتين فيدعى بالقيمة الحرجة .

فإذا وقع الإحصاء X فى المنطقة الحرجة فإننا نرفض H_0 لصالح H_1 ، أما إذا وقع X فى منطقة القبول فإننا نقبل H_0 .

نلاحظ أن احتمال ارتكاب خطأ من النوع الأول α ، وهو ما سميناه بمستوى معنوية الاختبار يحسب على النحو التالى :

$$\alpha = P \text{ (الخطأ من النوع الأول)}$$

$$= P\left(X \geq 7 | P = \frac{1}{5}\right)$$

$$= \sum_{x=7}^{15} b\left(x; 15, \frac{1}{5}\right)$$

$$= 1 - \sum_{x=0}^{6} b\left(x; 15, \frac{1}{5}\right)$$

$$= 1 - 0.9819$$

$$= 0.0181$$

وهكذا فإننا نختبر فرض العدم $P = \frac{1}{5}$ بمستوى معنوية قدره $(\alpha = 0.0181)$

ملاحظة :

يسمى مستوى المعنوية فى بعض الأحيان بحجم المنطقة الحرجة . والنتيجة السابقة تعنى المنطقة الحرجة ذات الحجم $\alpha = 0.0181$ صغيرة جداً ، وهذا يعنى أنه من غير المحتمل فى أننا سنرتكب خطأ من النوع الأول I . ومن غير الممكن حساب احتمال ارتكاب خطأ من النوع الثانى II قبل أن نحدد الفرض البديل تماماً . لنفرض مثلاً أن الفرض البديل هو $P = \frac{1}{4}$ ، ففى هذه الحالة فإننا سنكون متمكنين من حساب .

$\beta = p$ (احتمال ارتكاب خطأ من النوع الثانى)

ومنه : $\beta = P\left(X < 7 \mid P = \frac{1}{4}\right)$

$$= \sum_{x=0}^{6} b\left(x; 15, \frac{1}{4}\right)$$

ومنه : $= 0.9434$

ويشير هذا الاحتمال الكبير جداً إلى حد ما إذا ما قورن بـ $\alpha = 0.0181$ إلى ضعف عملية الاختبار ، وهو يعنى أننا سنرفض اللقاح الجديد الذى يفوق اللقاح المستخدم حاليا فى فعاليته . نرغب فى كثير من الأحيان فى استخدام إجراءات اختبار يكون معها كلا من α و β صغيرين . وهذا ممكن بالطبع وهو بيد من يجرى عملية الاختبار السابقة .

لنفرض مثلاً أن $H_1 : P = 0.6$ فى هذه الحالة نلاحظ أن :

$$\beta = P\ (\text{الخطأ من النوع الثانى})$$

$$= P(X < 7 \mid P = 0.6)$$

$$= \sum_{x=0}^{6} b(x; 15, 0.6)$$

$$= 0.0951$$

وبهذا الاحتمال الصغير نرتكب خطأ من النوع الثانى (II) وهو غير محتمل إلى أبعد حد . ذلك لأننا سنرفض اللقاح الجديد فى الوقت الذى نعلم فيه أن فعاليته بعد مرور عامين على استخدامه هى 60% . ونلاحظ أنه عندما يقترب الفرض البديل من الواحد $H_1 : p \to 1$ ، فإن β تقترب من الصفر .

ملاحظات هامة :

إن عملية إنقاص β ممكن دوما ، وذلك بزيادة حجم المنطقة الحرجة لنرى على سبيل المثال ماذا يحدث لكل من α و β عندما نغير نقطتنا الحرجة إلى 6.5 . ولنختبر فرض العدم $P = \frac{1}{5}$ ضد الفرض البديل $P = 0.6$.

نلاحظ أن :

$$\alpha = \sum_{x=6}^{15} b\left(x; 15, \frac{1}{5}\right)$$

$$= 1 - \sum_{x=0}^{5} b\left(x; 15, \frac{1}{5}\right)$$

$$= 1 - 0.9389$$

$$= 0.0611$$

وأن :

$$\beta = \sum_{x=0}^{5} b\,(x; 15, 0.6)\ .$$

$$= 0.0338$$

وهكذا نلاحظ أننا انقصنا احتمال ارتكاب خطأ من النوع الثانى على حساب زيادة احتمال ارتكاب الخطأ من النوع الأول . لحسن الحظ أنه يمكن إنقاص كلا الاحتمالين بزيادة حجم العينة . لنعيد نفس التجربة وذلك من أجل $n = 75$ شخص . لنفرض أننا نرفض فرض العدم $p = \frac{1}{5}$ فيما إذا اجتاز 25 شخصا أو أكثر العاملين دون التقاطهم الفيروس ، ونقبل عندئذ الفرض البديل $p > \frac{1}{5}$. إن النقطة الحرجة فى هذه الحالة هى 24.5 ، وجميع النقاط المحرزة والواقعة أعلى هذه النقطة تؤلف ما سميناه بالمنطقة الحرجة . كما أن جميع النقاط المحرزة التى تقع أدناها تشكل منقطة القبول .

سنستخدم المنحنى الطبيعى لحساب احتمال ارتكاب خطأ من النوع الأول (I) . وفى هذه الحالة نلاحظ أن :

$$\mu = np = (75)\left(\frac{1}{5}\right) = 15$$

$$\sigma = \sqrt{npq} = \sqrt{(75)\left(\frac{1}{5}\right)\left(\frac{4}{5}\right)} = 3.464$$

ويوضح القسم المظلل على الشكل (١,٧) هذا الاحتمال .

$\alpha = P$ (ارتكاب خطأ من النوع الأول)

$\alpha = P(X > 24.5 |$ فرض العدم صحيح)

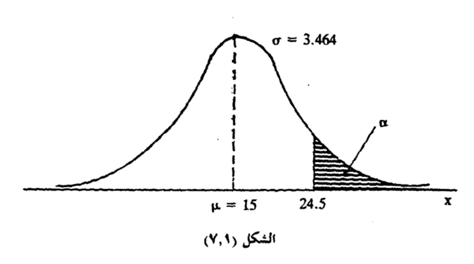
والقيمة Z المقابلة لـ $x = 24.5$ هى :

$$Z = \frac{24.5 - 15}{3.464} = 2.742$$

لذلك فإن :

$$\begin{aligned} \alpha &= P(Z > 2.742) \\ &= 1 - P(Z < 2.742) \\ &= 1 - 0.9969 \\ &= 0.0031 \end{aligned}$$

يمكن أن نحدد احتمال ارتكاب خطأ من النوع الثانى ، وذلك باستخدام المنحنى الطبيعى فيما إذا كان فرض العدم خاطىء والبديل صحيح أى أن $P = 0.6$ نلاحظ أن :



الشكل (٧,١)

$$\mu = np = (75)(0.6) = 45$$

$$\sigma = \sqrt{npq} = \sqrt{(75)(0.6)(0.4)} = 4.242$$

ويمثل القسم المظلل من الشكل (٧,٢) احتمال الوقوع ضمن منطقة القبول عندما يكون H_1 صحيحاً لذلك فإن :

β = p(ارتكـــاب خطـــأ من النـــــوع الثـــــانى)

= P (X < 24.5 | عندما يكون H صحيحاً)

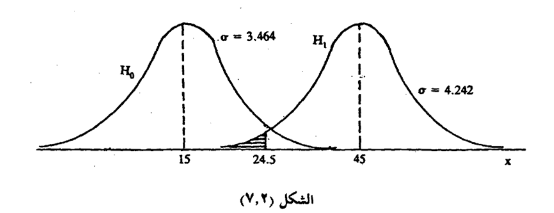
وتكون قيمة Z المقابلة للقيمة $x = 24.5$ هى :

$$Z = \frac{24.5 - 45}{4.242} = -4.83$$

لذلك فإن :

$$\beta = P(Z < -4.83)$$
$$\approx 0$$

والنتائج السابقة تشير إلى أنه قلما يحدث خطأ من النوع الأول أو النوع الثانى عندما تضم التجربة ٧٥ شخصاً .



الشكل (٧,٢)

مثال (٧,١)

أوجد كلا من α و β فى الاختبار :

$$H_0 \quad : \quad P = \frac{1}{4}$$

$$H_1 \quad : \quad P = \frac{1}{2}$$

وذلك عندما يكون للإحصاء X توزيعاً حدانياً فى تجربة كررت $n = 10$ مرات ، وبفرض أن المنطقة الحرجة معرفة بالعلاقة $X \geq 4$.

الحل

إن احتمال الرفض الخاطىء α هو :

$$\alpha = P\left[X \geq 4 \mid P = \frac{1}{4} \right]$$

$$= \sum_{x=4}^{10} \binom{10}{x} \left(\frac{1}{4}\right)^{x} \left(\frac{3}{4}\right)^{10-x}$$

$$= 0.2241$$

كما أن احتمال القبول الخاطىء β فيحسب بالعلاقة :

$$\beta = P\left(X < 4 | P = \frac{1}{2} \right)$$

$$= \sum_{x=0}^{3} \binom{10}{x} \left(\frac{1}{2}\right)^{x} \left(\frac{3}{4}\right)^{10-x}$$

$$= 0.1719$$

مثال (٧,٢)

بفرض أن $\alpha = 0.05$ فى المثال (٧,١) برهن أن المنطقة الحرجة هى $x \geq 6$.

الحل

نلاحظ أن :

$$\alpha = 0.05 = P\left[X \geq y | P = \frac{1}{4} \right]$$

$$= \sum_{x=y}^{10} \binom{10}{x} \left(\frac{1}{4}\right)^{x} \left(\frac{3}{4}\right)^{10-x}$$

ونلاحظ فى المثال (٧,١) أنه إذا كانت $y = 4$ لكانت $\alpha = 0.2281$. وبشكل مشابه إذا كانت $y = 5$ لكانت $\alpha = 0.0781$ ، وإذا كانت $y = 6$ لكانت $\alpha = 0.0197$. وبما أن y عدد صحيح ، لذلك ليس هناك قيمة لـ y تعطى قيمة لـ $\alpha = 0.05$ بالضبط ، فإذا كانت $y = 5$ عندئذ تكون α كبيرة جداً . أما إذا كانت $y = 6$ لكانت α صغيرة جدا . لذلك إذا أخذنا $y = 6$ فإننا نجد أن α لن تكون أكبر من 0.05 .

مثال (٧,٣)

لنفرض أننا نرغب فى تقدير وسط الإنتاج اليومى فى مصنع ، ولنفرض أننا سجلنا الإنتاج اليومى لفترة $n = 50$ يوما فكان وسط هذه العينة وانحرافها المعيارى بالأطنان كالتالى :

$$\bar{x} = 871 \quad , \quad S = 21$$

لنرمز بـ μ لوسط الإنتاج اليومى فى هذا المصنع . من المعلوم أن $\bar{x} = 871$ طناً يومياً هو التقدير الأفضل للوسط μ . لنختبر الفرض بأن معدل الإنتاج اليومى هو $\mu = 880$ يومياً ، والذى سنعتبره فرض عدم سنرمز له بالرمز H_0 أى أن :

$$H_0 \;:\; \mu = 880$$

ضد الفرض البديل بأن μ 880 طنا يومياً أى أن :

$$H_1 \;:\; \mu \neq 880$$

نلاحظ أن العينة المستخدمة $n = 50$ قياسا تحتوى على $S = 21, \bar{x} = 871$ ، وكما أوضحنا فإن الإحصاء $\bar{x}$ يمثل تقديراً نقطياً لـ μ أى أن إحصاء الاختبار هو :

$$Z = \frac{\bar{X} - H_0}{\sigma_{\bar{x}}} = \frac{\bar{X} - H_0}{\sigma / \sqrt{n}}$$

وباستخدام S كتقريب لـ σ نجد أن :

$$Z = \frac{871 - 880}{21 / \sqrt{50}} = -3.03$$

ومن أجل $\alpha = 0.05$ تكون منطقة الرفض $Z > 1.96$ أو $Z < -1.96$. وبما أن القيمة المحسوبة لـ Z تقع ضمن منطقة الرفض لذلك نرفض فرض العدم ونستنتج أن $\mu < 880$.

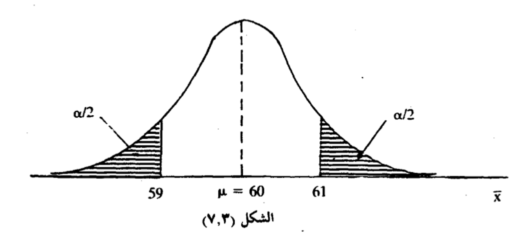
مثال (٧,٤)

لاختبار فرض العدم ، وهو أن وسط أوزان الطلاب فى كلية الهندسة هو 68 كيلوغراما ضد الفرض البديل فى أن هذا الوسط لا يساوى 68 كيلوغراما أى أن :

$$H_0 \;:\; \mu = 60$$
$$H_1 \;:\; \mu \neq 60$$

وذلك بفرض أن الانحراف المعيارى لمجتمع أوزان الطلاب فى هذه الكلية هو $\sigma = 3$ ، فإننا نختار عينة من طلاب كلية الهندسة بصورة عشوائية حجمها $n = 32$ طالبا . ثم نحدد قبل كل شيء الإحصاء $\bar{x}$ والذى سنبنى عليه قرارنا . نلاحظ أن الإحصاء $\bar{x}$ يمثل أفضل تقدير للوسط μ ، وأن هذا الإحصاء توزيع معاينة قريباً من التوزيع الطبيعى بالانحراف المعيارى $\sigma_{\bar{x}} = \sigma / \sqrt{n} = 0.53$.

لنحدد بعد ذلك النقطة الحرجة والمنطقة الحرجة ومنطقة القبول . نلاحظ أن وسط العينة $\bar{X}$ سيقع بالقرب من القيمة المختبرة 60 . لذلك نأخذ العددين 59 , 61 ثم نعتبر أن المنطقة الحرجة هى المنطقة $\bar{X} < 59$ و $\bar{X} > 61$. انظر الشكل (٧,٣) .



الشكل (٧,٣)

لذلك فإن منطقة القبول ستكون المنطقة $59 < \bar{X} < 61$. وهكذا نستنتج أنه إذا وقع وسط العينة $\bar{x}$ داخل القسم المظلل (المنطقة الحرجة) فإننا نرفض H_0 ، أما إذا وقع فى المجال (61 , 59) فإننا نقبلها .

لنحسب احتمال ارتكاب خطأ من النوع الأول ، نعلم أن :

$\alpha = P(\bar{X} < 59 |$ الفرض H_0 صحيح $) + P(\bar{X} > 61 |$ الفرض H_0 صحيح $)$

وقيم Z الموافقة للقيمتين $\bar{x}_1 = 59$, $\bar{x}_2 = 61$ عندما يكون H_0 صحيحاً هى القيم :

$$Z_1 \frac{59 - 60}{0.53} = -1.886$$

$$Z_2 \frac{51 - 60}{0.53} = 1.886$$

لذلك فإن :

$$\alpha = P(Z < -1.886 + P(Z > 1.886)$$
$$\alpha = 2P(Z < -1.886)$$

ومنه :

$$\alpha = 0.0588$$

هذا يعنى أن 0.06% من العينات ذات الحجم $n = 30$ ستقودنا إلى رفض الفرض $\mu = 60$ كيلوغراما فى الوقت التى تكون فيه هذه الفرض صحيحة . نلاحظ أنه لإنقاص α ، علينا أن نزيد حجم العينة ، أو أن نجعل منطقة القبول أوسع مما هى عليه الآن . لنفرض مثلا أن $n = 50$ نلاحظ أن :

$$\sigma_{\bar{x}} = 30 / \sqrt{50} = 0.424$$

كما أن :

$$Z_1 = \frac{59 - 60}{0.424} = -2.358$$

$$Z_2 = \frac{61 - 60}{0.424} = 2.358$$

لذلك فإن :

$$\alpha = P(Z < -2.358 + P(Z > 2.358)$$
$$= 2P(Z < -2.358)$$

ومنه :

$$= 0.0182$$

ولا يكفى إنقاص قيمة α لكى نضمن إجراءات الاختبار الذى نجربه جيدة . وعلينا تحديد قيمة β من أجل مختلف الفروض البديلة ، والتى نشعر أنها ستكون مقبولة إذا كانت صحيحة . لذلك فإن من الضرورى أن نرفض H_0 عندما يكون الوسط مساويا لإحدى القيم $\mu \geq 62$ أو $\mu \leq 58$. وفى هذه الحالة علينا أن نحدد احتمال ارتكاب خطأ من النوع الثانى β وفحصه ، وذلك من أجل الفرضين $\mu = 58$, $\mu = 62$.

ويكفى فقط (وذلك بسبب التناظر) أن نفكر فى احتمال قبول فرض العدم $\mu = 60$ ، وذلك عندما يكون الفرض البديل $\mu = 62$ صحيحاً . فإذا وقع وسط العينة $\bar{x}$ بين 59 و 61 عندما يكون H_1 صحيحاً . فإننا نكون قد ارتكبنا خطأ من النوع II . بالعودة إلى الشكل (٧,٤) نجد أن :

$$\beta = P(59 < \bar{X} < 61 \mid \text{صحيحاً } H_1 \text{ عندما يكون})$$

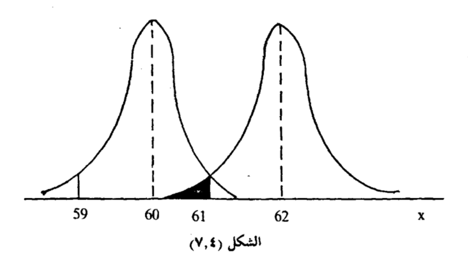
نلاحظ أن قيم Z المقابلة لـ $\bar{x}_1 = 59$, $\bar{x}_2 = 61$ عندما يكون H_1 صحيحاً هى :

$$Z_1 = \frac{59 - 62}{0.424} = -7.07$$

$$Z_2 = \frac{61 - 62}{0.424} = -2.36$$

لذلك فإن :

$$\beta = P(-7.07 < Z < -2.36)$$
$$= P(Z < -2.36) - P(Z < -7.07)$$
$$= 0.0091 - 0.0000$$
$$= 0.0091$$



الشكل (٧,٤)

إذا كانت القيمة الحقيقية لـ μ هى القيمة البديلة $\mu = 58$ ، فعندئذ ستكون قيمة β هى نفس العدد 0.0132 . من أجل جميع القيم $\mu < 58$ أو $\mu > 62$ ، فإن قيمة β ستكون صغيرة وذلك من أجل $n = 58$. لذلك فإنه سيطرأ تغير طفيف حول قبول H_0 فى الوقت الذى يكون فيه خاطئاً .

إن احتمال ارتكاب خطأ من النوع II سيزداد بسرعة عندما تقترب القيمة الحقيقية لـ μ من القيمة المفروضة دون أن تساويها ، وهذه الحالة عادية بالطبع وذلك عندما لا نفكر فى ارتكاب خطأ من النوع II فمثلا إذا كان الفرض البديل $\mu = 60.5$ صحيحاً ، فإننا لا نفكر فى ارتكاب خطأ من النوع II وذلك باستنتاج أن الإجابة الصحيحة هى μ

$\mu = 60$، واحتمال حصول مثل هذا الخطأ سيكون عاليا جدا عندما تكون $n = 58$.
بالعودة إلى الشكل (٧,٥) نجد أن :

$$\beta = P(59 < \bar{x} < 61 \mid \text{عندما يكون } H_1 \text{ صحيحاً})$$

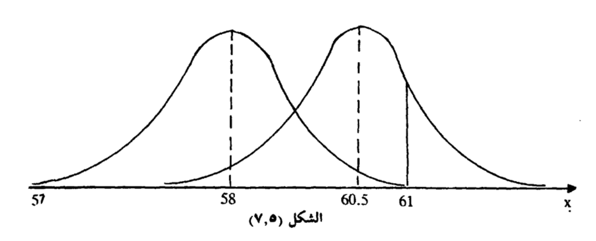
وقيم Z المقابلة للقيم $\bar{x}_1 = 59$ و $\bar{x}_2 = 61$ من أجل $\mu = 60.5$

$$Z_1 = \frac{59 - 60.5}{0.424} = -3.54$$

$$Z_2 = \frac{61 - 60.5}{0.424} = 1.18$$

لذلك فإن :

$$\beta = P(-3.54 < Z < -1.18)$$
$$= P(Z < 1.18) - P(Z < -3.54)$$
$$= 0.8810 - 0.0000$$
$$= 0.8810$$



الشكل (٧,٥)

توضح الأمثلة السابقة الخواص الهامة التالية :

(١) الخطأ من النوع I والخطأ من النوع II مرتبطان ببعضهما ، فأى زيادة فى احتمال أحدهما يؤدى إلى نقصان فى احتمال الآخر .

(٢) إن حجم المنطقة الحرجة ، واحتمال ارتكاب خطأ من النوع الأول يمكن

انقاصهما بتعديل القيمة الحرجة .

(٣) إن أى زيادة فى حجم العينة n سينقص كلا من α و β فى وقت واحد .

(٤) إذا كان فرض العدم خاطئاً ، فعندئذ يكون β أعظميا عندما تكون القيمة الحقيقية للمتغير قريبة (مجاورة) للقيمة المختبرة . تقابل المسافة الكبيرة بين القيمة الحقيقية والقيمة المختبرة قيمة صغيرة لـ β .

(٧,٣) الاختبارات وحيدة وثنائية الذيل One-tailed and two-tailed tests

نسمى اختبار فرض إحصائى يكون فيه فرض البديل ذو طرف واحد مثل :

$$H_0 : \theta = \theta_0$$
$$H_1 : \theta > \theta_0$$

أو :

$$H_0 : \theta = \theta_0$$
$$H_1 : \theta < \theta_0$$

باختبار وحيد الذيل . وتكون المنطقة الحرجة للفرض البديل $\theta > \theta_0$ تماما فى الذيل الأيمن للتوزيع ، بينما تكون المنطقة الحرجة للفرض البديل $\theta < \theta_0$ فى الذيل الأيسر تماما . يوضح المثال (٧,٣) الاختبار الوحيد الذيل .

سنسمى الاختبار الذى يكون فيه الفرض البديل ذو طرفين مثل :

$$H_0 : \theta = \theta_0$$
$$H_1 : \theta \neq \theta_0$$

باختبار ثنائى الذيل . تؤلف قيم الفرض البديل فى الذيلين ما يسمى بالمنطقة الحرجة . وكمثال على الاختبار ثنائى الذيل نذكر المثال (٧,٤) الذى سقناه حول اختبار وسط أوزان طلاب كلية الهندسة . وكنا قد فرضنا أن فرض العدم هو $\mu = 60$ كيلوغراما . كما أن الفرض البديل ثنائى الذيل هو $\mu \neq 60$.

ملاحظة هامة

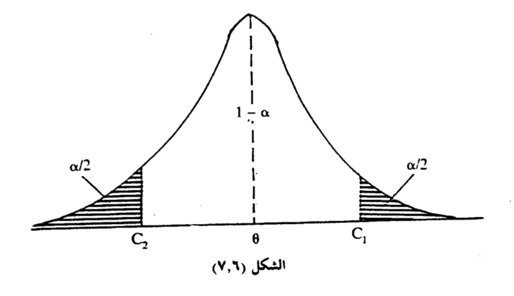
عند اختبار فرض يتعلق بمجتمع منقطع فإننا نقوم بشكل كيفي باختبار المنطقة الحرجة وتحديد حجمها . فإذا كان حجم هذه المنطقة α كبيراً جداً ، فإن بإمكاننا

انقاصه بتعديل موضع القيمة الحرجة . أما عندما نريد اختبار فرضية حول مجتمع مستمر ، فإننا نقوم باختبار قيمة α لتكون مساوية لـ 0.05 أو 0.01 على الغالب ، وبعدها نجد القيمة الحرجة C الموافقة . فمثلا فى الاختبار الثنائى الذيل بمستوى المعنوية $\alpha = 0.05$ من الأهمية . فإن القيمتين الحرجتين للإحصاء ذى التوزيع الطبيعى المعيارى ستكونان $-1.96 = Z_{0.025}$ ، وإذا استخدمنا مصطلح قيم Z فستكونان عندئذ المنطقة $Z_{-0.025} = 1.96$, $Z_{0.025} = -1.96$ الحرجة من الشكل $Z < -1.96$ و $Z > 1.96$. أما منطقة القبول فهى $-1.96 < Z < 1.96$.

نقول بأن اختبارنا فعالا إذا رفضنا H_0 تحت مستوى من المعنوية قدره $\alpha = 0.05$ ، ويعتبر هذا الاختبار أكثر فعالية إذا تم رفض H_0 تحت مستوى من المعنوية قدره $\alpha = 0.01$. أما إذا كنا نرفض H_0 لمجرد ألا يكون مساويا لـ θ أى سواءاً كان θ_0 أكبر أم أصغر من θ فعندما نعبر عن فرض العدم والبديل بالشكل التالى :

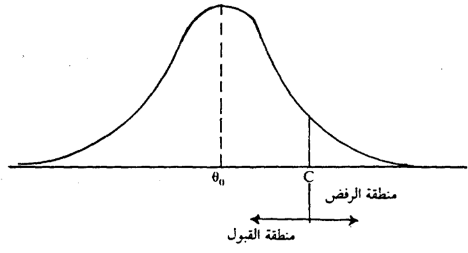
$$H_0 : \theta = \theta_0$$
$$H_1 : \theta \neq \theta_0$$

ويمكن القول بأن أفضل اختبار يتوفر لنا عملياً هو الاختبار الذى تتوضع فيه منطقة الرفض على التساوى فى ذيلى المنحنى . نلاحظ أن اختبارنا السابق هو ثنائى الذيل . وعندما نقتطع مساحة تساوى $\frac{\alpha}{2}$ من كل ذيل من ذيلى المنحنى ، فإننا نقوم عندئذ بتحديد نقطتين C_1 و C_2 بحيث يكون $P(\hat{\theta} \geq C_1 \mid H_0 \text{ صحيح}) = \frac{\alpha}{2}$ و $P(\hat{\theta} \leq C_1 \mid H_0 \text{ صحيح}) = \frac{\alpha}{2}$ هذا ونرفض الفرض H_0 إذا كانت قيمة $\hat{\theta}$ إلى يمين C_1 أو إلى يسار C_2 ونقبله فيما عدا ذلك . يوضح الشكل (٧,٦) شكل منطقة الرفض وتوزيع $\hat{\theta}$ عندما يكون H_0 صحيحاً .



الشكل (٧,٦)

لنفرض أننا نرغب فى اختبار فرض يتعلق بوسيط θ وأن التقدير النقطى $\hat{\theta}$ لهذا الوسيط له توزيع طبيعى بوسط يساوى $\hat{\theta}$ وإنحراف معياري σ_{θ}. فإذا كان فرض العدم $H_0 : \theta = \theta_0$ فعندئذ سيتوزع التقدير $\hat{\theta}$ وفقا للتوزيع الطبيعى بوسط قدره θ_0 ، وذلك كما هو واضح على الشكل (٧,٧) .



الشكل (٧,٧)

لنفرض أننا نهتم برفض الفرض H_0 عندما يكون $\theta > \theta_0$ ، فعندئذ يكون الفرض البديل $H_1 : \theta > \theta_0$ ويمكن البرهان على أن أفضل اختبار هو ذلك الذى يرفض H_0 عندما يكون $\hat{\theta}$ كبير . والمقصود بكبر $\hat{\theta}$ هو أن تكون المسافة بينه وبين θ_0 مساوية لعدد من الانحرافات المعيارية $\sigma_{\hat{\theta}}$. يوضح الشكل (٧,٧) موقع منطقتى الرفض والقبول . نسمى القيمة C التى تفصل منطقة الرفض عن منطقة القبول بالقيمة الحرجة لإحصاء الاختبار . هذا ويحدد مستوى المعنوية الاختبار α موقع C لأن احتمال أن يكون $\hat{\theta} \geq C$ علما أن H_0 صحيح (أى احتمال رفض H_0 على أنه صحيح) يساوى α ونعبر عن ذلك بالعلاقة $P(\hat{\theta} \geq C \mid H_0 \text{ صحيح}) = \alpha$. وعلى سبيل المثال إذا كان $\alpha = 0.05$، فإن $C = \theta_0 + 1.645\,\sigma_{\hat{\theta}}$. ونلاحظ أن منطقة الرفض تقع فى الذيل الأيمن للمنحنى ، لذلك فاختبارنا هذا هو اختبار وحيد الذيل .

(٧,٤) اختبار وسط وتباين مجتمع إحصائى Tests concerning means and variances

نفرض أن μ يمثل وسط مجتمع إحصائى تباينه σ^2 معلوم . لنختبر فرض العدم $H_0 : \mu = \mu_0$ ضد الفرض البديل ثنائى الذيل $H_1 : \mu \neq \mu_0$. وكتقدير مناسب لهذا

الوسيط نأخذ الإحصاء $\overline{X}$ ونبنى عليه قرارنا . من المعلوم كما ورد فى الفصل الخامس أن للمتغير $\overline{X}$ توزيعاً قريبا من التوزيع الطبيعى بالوسط : $\mu_{\overline{x}} = \mu$ والتباين $\sigma^2_{\overline{x}} = \frac{\sigma^2}{n}$ (يمثل هذا العدد n حجم عينة مختارة بصورة عشوائية من المجتمع المدروس) . إذا استخدمنا كمستوى معنوية لهذا الاختبار العدد α ، فعندئذ يمكن أن نجد عددين موجبين $\overline{x}_2$ و $\overline{x}_1$ بحيث يكون $\overline{x}_1 < \overline{X} < \overline{x}_2$ يحددان منطقة القبول . ويؤلف ذيلى التوزيع $\overline{X} < \overline{x}_1$ و $\overline{X} > \overline{x}_2$ ما يسمى بالمنطقة الحرجة .

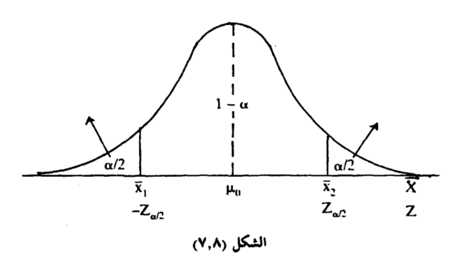
يمكن أن نعرف المنطقة الحرجة باستخدام قيم Z ، وذلك باستخدام علاقة التحويل :

$$Z = \frac{\overline{X} - \mu_0}{\sigma / \sqrt{n}}$$

لذلك ومن أجل مستوى المعنوية α ، فإنه يمكن مشاهدة القيم الحرجة للمتغير Z والموافقة للقيم $\overline{x}_2$ و $\overline{x}_1$ على الشكل (٧,٨) ، حيث إن :

$$-Z_{\alpha/2} = \frac{\overline{x}_1 - \mu_0}{\sigma / \sqrt{n}}$$

$$Z_{\alpha/2} = \frac{\overline{x}_2 - \mu_0}{\sigma / \sqrt{n}}$$



الشكل (٧,٨)

فإذا وقعت $\bar{x}$ فى منطقة القبول $\bar{x}_1 < \bar{X} < \bar{x}_2$ ، فعندئذ ستقع $Z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$ ضمن المنطقة $-Z_{a/2} < Z < Z_{a/2}$. نستنتج أن $\mu = \mu_0$. وهذا يعنى أننا سنرفض H_0 ونقبل الفرض البديل $\mu \neq \mu_0$.

إن إجراءات الاختبار التى درسناها سابقا تكافىء إيجاد $(1 - \alpha)100\%$ مجال ثقة للوسيط μ وقبول H_0 إذا كانت H_0 منتمية للمجال السابق . أما إذا رفضت μ_0 خارج المجال فإننا سنرفض الفرض H_0 لصالح الفرض البديل H_1 . وهكذا فإنه عندما يضع الدارس استنتاجات ما حول وسط μ لمجتمع تباينه σ^2 معلوم سواءا أكانت بإنشاء مجال ثقة أو عن طريق اختبار الفرضيات ، فإنه يستخدم نفس الإحصاء .

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$$

وبشكل عام إذا استخدم الدارس إحصاء ما لإنشاء مجال ثقة للمتغير θ سواءا أكان الإحصاء Z ، T ، χ^2 أو F ، فإن بإمكان استخدام نفس الإحصاء لاختبار الفرض فى أن المتغير (الوسيط) θ يساوى قيمة محددة θ_0 ضد أى فرض بديل . وبالطبع فإننا سنطبق جميع الإفتراضات الأساسية التى وردت فى الفصل السادس والمتعلقة باستخدام إحصاء ما ، وذلك لوصف الاختبارات المعروضة فى هذا الفصل . ويعنى هذا بشكل أساسى أننا نسحب عيناتنا من مجتمعات لها توزيعات قريبة من التوزيع الطبيعى . ومع ذلك فإن الإحصاء Z يمكن استخدامه لاختبار فرض حول وسطاء مجتمعات غير طبيعية فيها $n \geq 30$.

ويوضح الجدول (٧,٢) الطرق المستخدمة لاختبار فرض معين H_0 . كما يعطى المناطق الحرجة المناسبة للفروض وحيدة أو ثنائية الذيل . يمكن أن نجمل الخطوات المتعبة فى اختبار فرض متعلق بوسيط مجتمع θ ضد فرض بديل بالخطوات التالية :

نحدد أولا فرض العدم $H_0 : \theta = \theta_0$ ، وبعد ذلك نحدد الفرض البديل $H_1 : \theta < \theta_0$: أو $\theta > \theta_0$ أو $\theta \neq \theta_0$ ثم نختار مستوى المعنوية α ونحدد الاختبار الإحصائى المناسب ، وبعد ذلك المنطقة الحرجة . نحسب أيضا قيمة إحصاء الاختبار ، وذلك من خلال عينة حجمها n . وأخيرا فإننا نتخذ قراراً حول رفض H_0 إذا وقعت قيمة الإحصاء المحسوب فى المنطقة الحرجة ونقبلها فى عكس ذلك .

H_0	إحصاء الاختبار	H_1	المنطقة الحرجة
		$\mu < \mu_0$	$Z < -Z_\alpha$
$\mu = \mu_0$	$Z = \dfrac{\bar{X} - \mu_0}{\sigma / \sqrt{n}}$	$\mu > \mu_0$	$Z > Z_\alpha$
		$\mu \neq \mu_0$	$Z < -Z_{\alpha/2}$
			$Z > Z_{\alpha/2}$
		$\mu < \mu_0$	$T < -t_\alpha$
$\mu = \mu_0$	$T = \dfrac{\bar{X} - \mu_0}{S / \sqrt{n}}$; $v = n - 1$	$\mu > \mu_0$	$T > t_\alpha$
		$\mu \neq \mu_0$	$T > -t_{\alpha/2}$
		$\mu \neq \mu_0$	$T > t_{\alpha/2}$
		$\mu_1 - \mu_2 < d_0$	$Z < -Z_\alpha$
$\mu_1 - \mu_2 = d_0$	$Z = \dfrac{(\bar{X}_1 - \bar{X}_2) - d_0}{\sqrt{(\sigma_1^2/n_1) + (\sigma_1^2/n_2)}}$	$\mu_1 - \mu_2 > d_0$	$Z > Z_\alpha$
		$\mu_1 - \mu_2 \neq d_0$	$Z < -Z_{\alpha/2}$
			$Z > Z_{\alpha/2}$
		$\mu_1 - \mu_2 < d_0$	$T < -t_\alpha$
$\mu_1 - \mu_2 = d_0$	$T = \dfrac{(\bar{X}_1 - \bar{X}_2) - d_0}{S_p \sqrt{(1/n_1) + (1/n_2)}}$	$\mu_1 - \mu_2 > d_0$	$T > t_\alpha$
	$v = n_1 + n_2 - 2,\ \sigma_1 = \sigma_2$	$\mu_1 - \mu_2 \neq d_0$	$T < -t_{\alpha/2}$
	غير أنهما غير معلومين		$T > t_{\alpha/2}$
	$S_p^2 = \dfrac{(n_1 - 1) S_1^2 + (n_2 - 1) S_2^2}{n_1 + n_2 - 2}$		
$\mu_1 - \mu_2 = d_0$	$T' = \dfrac{(\bar{X}_1 - \bar{X}_2) - d_0}{\sqrt{(S_1^2/n_1) + (S_2^2/n_2)}}$;	$\mu_1 - \mu_2 < d_0$	$T' < -t_\alpha$
		$\mu_1 - \mu_2 > d_0$	$T' > t_\alpha$
		$\mu_1 - \mu_2 \neq d_0$	$T' < -t_{\alpha/2}$,
			$T' > t_{\alpha/2}$

H_0	إحصاء الاختبار	H_1	المنطقة الحرجة
	$S^2 = \dfrac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$ $\nu = \dfrac{(S_1^2/n_1 + S_2^2/n_2)^2}{\dfrac{(S_1^2/n_1)^2}{n_1 - 1} + \dfrac{(S_1^2/n_2)^2}{n_2 - 1}}$ وغير معلومين σ_2 σ_1		
			$\chi^2 < \chi_{1-\alpha}^2$
$\sigma^2 = \sigma_0^2$	$\chi^2 = \dfrac{(n-1)S^2}{\sigma_0^2}$	$\sigma^2 < \sigma_1^2$	
			$\chi^2 > \chi_\alpha^2$
		$\sigma^2 > \sigma_0^2$	$\chi^2 < \chi_{1-\alpha/2}^2$
		$\sigma^2 \neq \sigma_0^2$	$\chi^2 > \chi_{\alpha/2}^2$
			$F < f_{1-\alpha}(\nu_1, \nu_2)$
$\sigma_1^2 = \sigma_2^2$	$F = S_1^2/S_2^2$	$\sigma_1^2 < \sigma_2^2$	$F > f_\alpha(\nu_1, \nu_2)$
	$\nu_1 = n_1 - 1$,	$\sigma_1^2 > \sigma_2^2$	$F < f_{1-\alpha/2}(\nu_1, \nu_2)$,
	$\nu_2 = n_2 - 1$	$\sigma_1^2 \neq \sigma_2^2$	$F > f_{\alpha/2}(\nu_1, \nu_2)$

الجدول (٧,٢)

مثال (٧,٥)

من المعلوم أن واحدا من عشرة من المدخنين على وجه التقريب فى المملكة العربية السعودية يفضلون نوعا معينا من السجاير . وقد قامت الشركة المنتجة لهذا النوع بحملة دعائية واسعة لهذا النوع داخل المملكة وبعدها ، وبقصد اختبار فعالية هذه الحملة قامت الشركة بأخذ عينة مؤلفة من 200 مدخن فتبين لها أن 26 مدخناً منهم يفضلون هذا النوع . فهل كانت هذه الحملة ذات فائدة ؟

الحل

نلاحظ أن التجربة تحقق شروط التجربة الحدانية . نفرض أن p يمثل احتمال تفضيل

مدخن ما فى المملكة للنوع المذكور من السجاير ، ولنختبر فرض العدم .

$$H_0 : p = 0.10$$

ضد الفرض البديل :

$$H_1 : p > 0.10$$

ومن الطبيعى أن تكون p كذلك ، لأننا نهتم بالكشف عن أية زيادة فى الاحتمال p منطلقين من الاعتقاد بأن الحملة الدعائية إن لم تكن مفيدة فهى على أى حال لا يمكن أن تكون ضارة فتخفض من قيمة p . هذا وأن أفضل اختبار لمثل هذا الفرض هو الاختبار الناتج عن وضع كامل منطقة الرفض فى الذيل الأيمن لتوزيع إحصاء الاختبار . نعلم أن $\hat{p} = \frac{x}{n}$ يمثل تقديرا للنسبة p ، ويكون إحصاء الاختبار عندئذ $Z = \frac{\hat{p} - p_0}{\sigma_{\hat{p}}} = \frac{p - p_0}{\sqrt{p_0 q_0/n}}$.
وهذا يعنى أن $Z = \frac{x - np_0}{\sqrt{np_0q_0}}$ ، حيث فرضنا أن $P = P_0$ فى عبارة الانحراف المعيارى لـ p وهى $\sqrt{npq}$. وإذا اخترنا مستوى المعنوية $\alpha = 0.05$ ، فإننا نلاحظ من الجدول IV أن $Z = 1.645$.
وعلى هذا فإننا سنرفض H_0 عندما يكون $Z > 1.645$. وبحساب Z نجد أن :

$$Z = \frac{26 - 20}{\sqrt{200\,(0.1)\,(0.9)}} = 1.41$$

والقيمة الأخيرة للإحصاء لا تقع ضمن منطقة الرفض ، وبالتالى فإننا نقبل الفرض H_0 .
وهذا يعنى أن الحملة الدعائية كانت مفيدة حقا .

مثال (٧,٦)

بفرض أن وسط الزمن اللازم لتسجيل أى طالب بالنسبة لجميع الصفوف فى كلية الهندسة هو 50 دقيقة وأن الانحراف المعيارى هو 10 دقائق . استخدمنا طريقة حديثة أخرى للتسجيل ثم سحبنا عنية من الطلاب بصورة عشوائية مؤلفة من اثنى عشر طالباً ، فوجد أن لها وسط تسجيل قدره 42 دقيقة وانحرافاً معيارياً قدره 11.9 دقيقة . اختبر الفرض فى أن وسط المجتمع أقل من 50 ، وذلك باستخدام ١ – مستوى معنوية $\alpha = 0.05$ ٢ – ثم باستخدام مستوى معنوية $\alpha = 0.01$. افرض أن مجتمع الأزمنة طبيعى .

الحل

(١) نلاحظ أن فرض العدم هو دقيقة $H_0 : \mu = 50$

(٢) كما نلاحظ أن الفرض البديل هو دقيقة 50 > $H_1 : \mu$

(٣) وأن مستوى المعنوية فى الحالة الأولى 0.05 = α ، وفى الحالة الثانية $\alpha = 0.01$.

(٤) لتحديد المنطقة الحرجة نعود إلى الجدول (٧,٢) ونلاحظ أنه من أجل :

$$T = \frac{\overline{X} - \mu_0}{S / \sqrt{n}} : \nu = 11$$

أن $T < -1.796$ من أجل مستوى المعنوية $\alpha = 0.05$ وأن $T < -2.718$ من أجل $\alpha = 0.01$

(٥) لنحسب من خلال العينة ذات الحجم $n = 12$ قيمة الإحصاء T . نلاحظ أن $x = 42$ دقيقة ، $S = 11.9$ و $n = 12$ لذلك فإن :

$$t = \frac{42 - 50}{11.9 / \sqrt{12}} = -2.33$$

(٦) ثم نتخذ قرارنا التالى : نرفض الفرض H_0 بمستوى المعنوية 0.05 ونقبله من أجل مستوى المعنوية 0.01 . وهذا يعنى بشكل جوهرى أن من المرجح أن يكون الوسط الحقيقى لزمن التسجيل أقل من 50 دقيقة .

مثال (٧,٧)

طور مهندسو مصنع للتجهيزات الرياضية نوعا من الخيوط لصيد الأسماك ، وادعو أن لهذا النوع وسط قوة انقطاع قدره $\mu = 8$ كيلوغراماً ، وانحرافاً معيارياً قدره 0.5 كيلوغراماً . اختبر فرض العدم $H_0 : \mu = 8$ كيلوغراما ضد الفرض البديل $H_1 : \mu \quad 8$ كيلوغراما ، وذلك بفرض أن اخترنا عينة عشوائية مؤلفة من 50 خيطاً . وجربت فأعطت وسطا عينيا لقوة الانقطاع قدره 7.5 كيلوغراما . استخدم مستوى المعنوية $\alpha = 0.01$.

الحل

نلاحظ أن $H_0 : \mu = 8$ وأن $H_1 : \mu \neq 8$. أما بالنسبة لمستوى المعنوية فهو $\alpha = 0.01$. لذلك فإن

المنطقة الحرجة هى $Z < -2.58$ و $Z > 2.58$ حيث يمثل $Z = \frac{\overline{X} - \mu_0}{\sigma / \sqrt{n}}$. فإذا فرضنا أن $\overline{x} = 7.8$ و $n = 50$ فإننا نجد أن :

$$Z = \frac{7.8 - 8}{0.5 / \sqrt{50}} = -2.828$$

وحيث أن هذه القيمة تقع ضمن منطقة الرفض ، لذلك فإننا نرفض العدم H_0 ، ونستنتج أن وسط قوة الانقطاع ليس مساوياً لـ 8 كيلو غراما ، بل هو أقل من ثمانية كيلوغرامات .

مثال (٧,٨)

أجريت تجربة لمقارنة تآكل نوعين من المواد . فأخذت عينة مؤلفة من 12 قطعة من المادة الأولى واختبرت بتعريض كل قطعة من قطعها لآلة لقياس التآكل . كما اختبرت عينة أخرى من المادة الثانية مؤلفة من عشر قطع واختبرت بنفس الطريقة ولوحظ عمق التآكل فى كل حالة . وقد بلغ وسط التآكل بالنسبة لمادة العينة الأولى 85 وحدة وانحرافها المعيارى 4 ، بينما بلغ وسط التآكل بالنسبة لمادة العينة الثانية 81 وانحرافها المعيارى 5 . اختبر الفرض التالى : أن للمادتين السابقتين نفس التآكل بمستوى من المعنوية قدره $\alpha = 0.1$. وذلك إذا فرضنا أن لمجتمعى المادتين السابقتين توزيعاً قريباً من التوزع الطبيعى بتباينين متساويين .

الحل

نفرض أن μ_1 , μ_2 يمثلان وسطى مجتمعى المادتين الأولى والثانية على الترتيب ، لنتبع الخطوات الست التالية :

(١) $H_0 : \mu_1 = \mu_2$. أو $\mu_1 - \mu_2 = 0$

(٢) $H_1 : \mu_1 \neq \mu_2$. أو $\mu_1 - \mu_2 = 0$

(٣) $\alpha = 0.1$

(٤) المنطقة الحرجة هى $T < -1.725$ و $T > 1.725$ حيث يمثل :

$$T = \frac{(\overline{X}_1 - \overline{X}_2) - d_0}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \;; \quad = 20$$

(٥) بالحساب لدينا $n_1 = 12$ ، $S = 4$ ، $\bar{x}_1 = 85$ ، $n_2 = 10$ ، $S = 5$ ، $\bar{x}_2 = 81$ لذلك فإن :

$$S_p = \sqrt{\frac{(11)(16) + (9)(25)}{12 + 10 - 2}} = 4.478$$

$$t = \frac{(85 - 81) - 0}{4.478\sqrt{(1/12) + (1/10)}} = 2.07$$

(٦) الاستنتاج : بما أن $t = 2.07$ (قد وقعت فى منطقة الرفض) لذلك فإننا سنرفض فرض العدم H_0 ونستنتج أنه ليس للمادتين السابقتين نفس التآكل .

مثال (٧,٩)

استخدمت خمس عينات من مادة حديدية لتحديد ما إذا كان هناك فرقا بين التحليل الكيميائى المخبرى والتحليل الفلورى بأشعة X لمعرفة كمية الحديد الموجودة فى هذه المادة . نفرض أن كل عينة قد جزئت إلى عينتين جزئيتين ، وأنه قد طبقت على كل منهما طريقتى التحليل السابقتين . وقد أوضحت نتائج التحليل هذه كمية الحديد الموجودة بالجدول التالى :

العينة

التحليل	1	2	3	4	5
بأشعة X	2.0	2.0	2.3	2.1	2.4
بالتحليل الكيميائى	2.2	1.9	2.5	2.3	2.4

اختبر بمستوى من المعنوية قدره $\alpha = 0.05$ ما إذا كانت طريقتا التحليل المذكورتين تعطيان فى الوسط نفس النتائج . وذلك بفرض أن المجتمعات طبيعية .

الحل

نفرض أن وسط كمية الحديد المحددة بالطريقة الكيميائية هى μ_1 ، وبالطريقة الشعاعية هى μ_2 . نلاحظ باستخدام الخطوات الست المذكورة سابقا أن :

(١) $H_0 : \mu_1 = \mu_2,\ \mu_D \neq 0$

(٢) $H_1 : \mu_1 \neq \mu_2,\ \mu_D \neq 0$

(٣) $\alpha = 0.05$

(٤) بالعودة إلى السطر السادس فى الجدول (٧,٢) وبفرض أن $T = \frac{\bar{D} - d_0}{S / \sqrt{n}}$ بـ 4 = درجات من الحرية ، فإننا نلاحظ أن المنطقة الحرجة هى $T < -2.776$ و $T > 2.776$

(٥) الحسابات

التحليل الشعاعى	التحليل الكيميائى	d_1	d_1^2
2.0	2.2	−0.2	0.04
2.0	1.9	−0.1	0.01
2.3	2.5	−0.2	0.04
2.1	2.3	−0.2	0.04
2.4	2.4	0.0	0.00
		−0.5	**0.13**

ونجد أن $\bar{D} = -0.5/5 = -0.1$ وأن :

$$S_d = [(5)(0.13) - (0.5)^2] / (5)(4) = 0.2$$

وبأخذ الجذر التربيعى نجد أن $S_d = 0.14142$

لذلك فإن :

$$t = \frac{-0.1 - 0}{0.14142 / \sqrt{5}} = -1.6$$

(٦) الاستنتاج : بما أن القيمة $t = -1.6$ تقع ضمن منطقة القبول ، لذلك فإننا نقبل الفرض H_0 ونستنتج أنه لايوجد فرق جوهرى بين طريقتى التحليل المذكورتين .

مثال (٧,١٠)

يدعى صاحب للبطاريات أن عمر بطارياته له توزيعا قريبا من التوزيع الطبيعى بانحراف معيارى قدره 0.9 سنة . أخذت عينة عشوائية حجمها $n = 10$ من هذه البطاريات ، فوجد أن لها انحرافا معياريا عينيا قدره 1.2 سنة . فهل تعتقد أن $\sigma > 0.9$ سنة ؟ استخدم مستوى المعنوية $\alpha = 0.05$.

الحل

نلاحظ باتباع الخطوات الست المعروفة أن :

(١) $H_0 : \sigma^2 = 0.81$

(٢) $H_1 : \sigma^2 > 0.81$

(٣) $\alpha = 0.05$

(٤) من السطر السابع فى الجدول (٧,٢) نلاحظ وبفرض أن $\chi^2 = (n - 1) S^2/\sigma_0^2$ وحيث إن $\nu = 9$ ، أن المنطقة الحرجة فى هذه الحالة هى $\chi^2 > 16.919$

(٥) الحسابات : نلاحظ أن $n = 10 , S^2 = 1.44$ وأن :

$$\chi^2 = \frac{(9)\,(1.44)}{0.81} = 16.0$$

(٦) الاستنتاج : بما أن القيمة $\chi^2 = 16.0$ تقع ضمن منطقة القبول ، لذلك فإننا نقبل الفرض H_0 ونستنتج أنه لا يوجد أى سبب للشك فى أن $\sigma = 0.9$ سنة .

(٧,٥) اختيار حجم العينة لاختبار الوسط
Choice of sample size for testing mean

يمكن للمجرب أن يتحكم بمستوى معنوية اختبار إحصاء الفروض ، بينما يمكن التحكم بـ β أو قوة هذا الاختبار والذى نرمز له بالرمز $1 - \beta$ باستخدام حجم عينة مناسبة . سنناقش فى هذا الفصل طريقة اختبار حجم عينة للاختبارات التى تستلزم وسطاً أو (وسطين) لمجتمعاتها . نلاحظ أن مسألة تحديد حجم العينة الضرورى لتحقيق قوة اختبار معينة هى مسألة سهلة فى الحالة التى يكون فيها التوزيع المستخدم طبيعيا وتباينه (أو تبايناته) معلوما . لنفرض أننا نرغب فى اختبار الفرض $H_0 : \mu = \mu_0$ وذلك ضد الفرض البديل $H_1 : \mu > \mu_0$ بمستوى من المعنوية قدره α . وذلك بفرض أن تباين المجتمع المدروس σ^2 معلوما . فعلى سبيل المثال لنفرض أن الفرض البديل المحدد هو $\mu = \mu_0 + \delta$ عندئذ تعطى قوة هذا الاختبار بالعلاقة التالية :

$$1 - \beta = P\left[\frac{\overline{X} - \mu_0}{\sigma / \sqrt{n}} > Z_\alpha | H_1\right]$$

$$= P\left[\frac{\overline{X}-(\mu_0+\delta)}{\sigma/\sqrt{n}} > Z_\alpha - \frac{\delta}{\sigma/\sqrt{n}} \Big| \mu = \mu_0 + \delta\right]$$

وبالنسبة للفرض البديل $\mu = \mu_0 + \delta$ ، فإن للإحصاء $\frac{\overline{X}-(\mu_0+\delta)}{\sigma/\sqrt{n}}$ توزيعا طبيعياً معيارياً لذلك فإن :

$$1-\beta = P\left(Z > Z_\alpha - \frac{\delta\sqrt{n}}{\sigma}\right)$$

والتى نستنتج منها أن :

$$-Z_\beta = Z_\alpha - \frac{\delta\sqrt{n}}{\sigma}$$

ولذلك فإن :

$$n = \frac{(Z_\alpha + Z_\beta)^2\sigma^2}{\delta^2}$$

ملاحظة هامة

هذه النتيجة صحيحة فى الحالة التى يكون فيها الفرض البديل $\mu > \mu_0$. فى حالة اختبار ثنائى الذيل نحصل على قوة هذا الاختبار $1-\beta$ من أجل فرض بديل محدد عندما يكون :

$$n \simeq \frac{(Z_{\alpha/2} + Z_\beta)^2\sigma^2}{\delta^2}$$

مثال (٧,١١)

لاختبار فرض العدم $\mu_0 : \mu = 68$ ضد الفرض البديل $H_1 : \mu > 68$ بمستوى من المعنوية قدره $\alpha = 0.05$ ، وذلك بفرض أن $\sigma = 5$ ، فإننا نختار عينة من المجتمع المدروس

ونقوم بدراستها أوجد حجم العينة المطلوب إذا كانت قوة هذا الاختبار $1 - \beta = 0.95$ والوسط الحقيقى هو 69 .

الحل

نلاحظ أن $\alpha = \beta = 0.05$ ، وأن $Z_\alpha = Z_\beta = 1.645$. ومن أجل الفرض البديل $\mu = 69$ نأخذ $\delta = 1$ لذلك فإن :

$$n = \frac{(1.645 + 1.645)^2 (25)}{1} = 270.6$$

هذا يعنى أن الاختبار يتطلب 271 ملاحظة ، وذلك لرفض فرض العدم فى 95% من الحالات عندما تكون μ أكبر بكثير من 69 . يمكن استخدام نفس الإجراءات لتحديد حجم العينة $n = n_1 = n_2$ المطلوبة لدى القيام باختبار وسطى مجتمعين بقوة اختبار محددة . فمثلا لنفرض أننا نرغب فى اختبار فرض العدم $H_0 : \mu_1 : \mu_2$ ضد الفرض البديل $H_1 : \mu_1 \quad \mu_2$ عندما يكون σ_1 , σ_2 معلومين ، فمن أجل الفرض البديل المحدد $\mu_1 - \mu_2 = \delta$ تكون قوة اختبارنا محددة بالعلاقة التالية :

$$1 - \beta = P\left[\frac{\bar{X}_1 - \bar{X}_2}{\sqrt{(\sigma_1^2 + \sigma_2^2)/n}} > Z_{\alpha/2} | \mu_1 - \mu_2 = \delta\right]$$

لذلك فإن :

$$\beta = P\left[-Z_{\alpha/2} < \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{(\sigma_1^2 + \sigma_2^2)/n}} < Z_{\alpha/2} | \mu_1 - \mu_2 = \delta\right]$$

$$\beta = P\left[Z_{\alpha/2} - \frac{\delta}{\sqrt{\sigma_1^2 + \sigma_2^2)/n}} < \frac{\bar{X}_1 - \bar{X}_2 - \delta}{\sqrt{\sigma_1^2 + \sigma_2^2) / n}} < Z_{\alpha/2} \right.$$

$$\left. \frac{\delta}{\sqrt{(\sigma_1^2 + \sigma_2^2)/n}} | \mu_1 - \mu_2 = \delta \right]$$

وبالنسبة لهذا الفرض البديل $\delta = \mu_1 - \mu_2$ ، فإن للإحصاء $\frac{\bar{X}_1 - \bar{X}_2 - \delta}{\sqrt{(\sigma_1^2 + \sigma_2^2/n}}$ توزيعاً طبيعياً معيارياً لذلك فإن :

$$\beta = P\left[-Z_{\alpha/2} - \frac{\delta}{\sqrt{(\sigma_1^2 + \sigma_2^2)/n}} < Z < Z_{\alpha/2} - \frac{\delta}{\sqrt{(\sigma_1^2 + \sigma_2^2)/n}} \right]$$

من هذه العلاقة الأخيرة نستنتج أن :

$$-Z_\beta \simeq Z_{\alpha/2} - \frac{\delta}{\sqrt{(\sigma_1^2 + \sigma_2^2)/n}}$$

لذلك فإن :

$$n \simeq \frac{(Z_{\alpha/2} + Z_\beta)^2 (\sigma_1^2 + \sigma_2^2)}{\delta^2}$$

وعبارة حجم العينة المطلوبة من أجل اختبار وحيد الذيل عندما يكون $n = n_1 = n_2$ تعطى بالعلاقة :

$$n = \frac{(Z_{\alpha/2} + Z_\beta)^2 (\sigma_1^2 + \sigma_2^2)}{\delta^2}$$

(٧,٦) الاختبارات المتعلقة بالنسب
Tests concerning proportions

يتطلب فى كثير من مجالات الحياة إجراء اختبارات فروض تتعلق بالنسب ، فمثلا يهتم رجل السياسة عادة فى معرفة نسبة الأصوات المؤيدة له فى الانتخابات القادمة . كما تهتم جميع الشركات الصناعية بتحديد نسبة العناصر المعابة وذلك عند شحن كمية من البضائع المصنعة . سندرس مسألة اختبار فرض العدم $H_0 : p = p_0$ وذلك ضد الفرض البديل وحيد الذيل $H_1 : p < p_0$ أو $H_1 : p > p_0$ أوثنائى الذيل $H_1 : p \neq p_0$ وذلك بفرض أن p يمثل متغير (وسيط) التوزيع الحدانى .

إن الإحصاء المناسب والذى سنبنى عليه قرارنا فى رفض أو قبول فرض العدم H_0 هو المتغير العشوائى X . كما يمكن أن نأخذ الإحصاء $\hat{p} = \frac{X}{n}$. وستقودنا قيم المتغير X

والتى تكون بعيدة عن الوسط $\mu = np_0$ إلى رفض الفرض H_0 . ولاختبار الفرض $H_0 : p = p_0$ ضد الفرض البديل $H_1 : p < p_0$ فإننا نستخدم التوزيع الحدانى بالمتغير $p = p_0$ و $q = 1 - p_0$ لتحديد $P(X \leq x |$ صحيح $H_0)$ والقيمة x تمثل عد النجاحات فى عينتان والتى يساوى حجمها n . إذا كان $P(X \leq x |$ صحيح $H_0) < \alpha$ فيكون اختبارنا على مستوى من المعنوية قدره α ونرفض عندئذ H_0 لصالح H_1 . وبصورة مشابهة لاختبار الفرض $H_0 : p = p_0$ ضد الفرض البديل $H_1 : p > p_0$ ، فإننا نبحث عن $P(X \geq x |$ صحيح $H_0)$ ونرفض H_0 لصالح H_1 إذا كان هذا الاحتمال أقل من α . وأخيرا لاختبار الفرض $H_0 : p = p_0$ ضد الفرض البديل $H_1 : p \neq p_0$ بمستوى المعنوية α ، فإننا نرفض H_0 عندما يكون $P(X \leq x |$ صحيحة $H_0) < \alpha/2, x < np_0$ أو إذا كان $x > np_0$ و $P(X \geq x | H_0) < \frac{1}{2}$.
يمكن أن نجمل الخطوات المتبعة فى اختبار فرض حول النسبة ضد مختلف الفروض البديلة بما يلى :

(١) نحدد فرض العدم $H_0 : p = p_0$

(٢) نحدد الفرض البديل $p < p_0$ أو $p > p_0$ أو $H_1 : p \neq p_0$

(٣) نختار مستوى معنوية الاختبار α .

(٤) نحدد المنطقة الحرجة .

أ ـــ فمن أجل الفرض البديل $p < p_0$ تكون المنطقة الحرجة ممثلة لجميع قيم x المحققة للمتباينة $P(X \leq x |$ صحيح $H_0) < \alpha$

ب ـــ أما من أجل $H_1 : p > p_0$ فتؤلف مجموعة القيم x المحققة للمتباينة $P(X \geq x |$ صحيح $H_0) < \alpha$ ما يسمى بالمنطقة الحرجة فى مثل هذه الحالة .

جـ ـــ وأخيراً فإن جميع قيم x المحققة للمتباينة $P(X \leq x |$ صحيح $H_0) < \alpha/2$ وذلك من أجل $x < np_0$ أو جميع قيم x المحققة لـ $P(X \geq x |$ صحيح $H_0) < \alpha/2$ من أجل $x > np_0$ تؤلف منطقة حرجة للفرض البديل $H_1 : p \neq p_0$

(٥) نجد x ثم نحسب الاحتمال الموافق

(٦) نتخذ قرار حول رفض H_0 فيما إذا وقع x فى المنطقة الحرجة ونقبله فيما عدا ذلك .

مثال (٧,١٢)

يدعى صياد أنه يصيب 90% من الطيور التى يطلق عليها عياراته النارية . هل توافق الصياد إدعائه هذا إذا علمت أنه أصاب فى يوم من الأيام أثنى عشر طيراً عندما

أطلق عليها عشرون طلقة متتالية ؟ استخدم مستوى معنوية قدره $\alpha = 0.05$.

الحل

لحل هذا المثال نتبع الخطوات الست التالية :

(١) نلاحظ أن فرض العدم هى $H_0 : p = 0.9$

(٢) وأن الفرض البديل هى $H_1 : p \neq 0.9$

(٣) أما بالنسبة لمستوى المعنوية فهو $\alpha = 0.05$

(٤) نلاحظ أن جميع قيم x المحققة للمتباينة $P(X \leq x |$ صحيح $H_0) < 0.025$ تحدد المنطقة الحرجة .

(٥) كما أن لدينا $x = 12$ وأن $n = 20$. لذلك وباستخدام الجدول II نجد أن :

$$P(X \leq 12 | p = 0.9) = \sum_{x=0}^{12} b(x;20\,,\,0.9)$$

$$= 0.0004 < 0.025$$

(٦) نرفض الفرض H_0 ونستنتج أن ادعاء هذا الصياد خاطىء .

وجدنا فى الفقرة (٣,٤) أن الاحتمال الحدانى قد استنتج صيغة التوزيع الحدانى الفعلى أو بواسطة الجدول II ، وذلك فى الحالة التى تكون فيها n صغيرة . أما إذا كانت n كبيرة فإننا نجرى عمليات التقريب المطلوبة . أما عندما تكون القيمة المختبرة p_0 قريبة من الصفر أو الواحد ، فإننا نستخدم التوزيع البواسونى بالوسط $\mu = np$. هذا ويقدم المنحنى الطبيعى تقريباً يكون عادة مفضلا فى الحالة التى تكون فيها n كبيرة و p_0 ليست قريبة من الصفر ولا الواحد . باستخدام التقريب الطبيعى فإننا نبنى قرارنا على المتغير الطبيعى المعيارى :

$$Z = \frac{\hat{P} - p_0}{\sqrt{p_0\, q_0 / n}} = \frac{X - np_0}{\sqrt{np_0\, q_0}}$$

لذلك فإن المنطقة الحرجة للاختبار ثنائى الذيل بمستوى المعنوية α هى المنطقة $Z > Z_{\alpha/2}$ ،

$Z < -Z_{\alpha/2}$. أما بالنسبة للاختبار وحيد الذيل من أجل الفرض البديل $p < p_0$ فتكون المنطقة $Z < -Z_\alpha$ ممثلة للمنطقة الحرجة فى هذا الاختبار . وأخيرا فبالنسبة للفرض البديل $p > p_0$ تكون المنطقة الحرجة هى $Z > Z_\alpha$.

لاختبار فرض يتعلق بالنسبة وذلك باستخدام منحنى التقريب الطبيعى ، فإننا نتبع الخطوات التالية :

(١) فرض العدم $H_0 : p = p_0$
(٢) الفرض البديل $p > p_0$ $p < p_0$ أو $H_1 : p \neq p_0$
(٣) نختار مستوى المعنوية الاختبار α .
(٤) نحدد المنطقة الحرجة .
أ — $Z < -Z_\alpha$ إذا كانت الفرض البديل $p < p_0$
ب — $Z > Z_\alpha$ إذا كانت الفرض البديل $p > p_0$
جـ — $Z < -Z_{\alpha/2}$ وذلك من أجل الفرض البديل $p \neq p_0$
(٥) نحسب x من عينة حجمها n وبعد ذلك نحسب $Z = \dfrac{x - np_0}{\sqrt{np_0 q_0}}$
(٦) نتخذ قراراً في رفض H_0 إذا وقع Z في المنطقة الحرجة ونقبله فيما عدا ذلك .

مثال (٧,١٣)

ادعت شركة صناعية أن 95% من عناصرها المنتجة هى جيدة الصنع . وقد قامت الشركة مؤخراً باستخدام وسائل جديدة ومتطورة بقصد خفض نسبة الخمسة بالمائة من العناصر المعابة فى انتاجها . ثم أنتجت فى يوما ما مائة عنصر ، وذلك باستخدام هذه الطرق الحديثة . بعد ذلك قامت بفحصها فوجدت أن ثلاثة عناصر من بينها معابة . فهل تكون هذه العينة دليلا كافيا لاستنتاج أن الوسائل الحديثة قد طورت فعلا عملية الإنتاج هذه ؟ استخدم 0.01 مستوى معنوية .

الحل

نلاحظ باستخدام الخطوات الست المذكورة سابقاً أن :

(١) $H_0 : P = 0.95$
(٢) $H_1 : p > 0.95$

(٣) $\alpha = 0.01$

(٤) المنطقة الحرجة هى $Z > 2.575$

(٥) لدينا $x = 97$ و $n = 100$ لذلك فإن $npo = (100) - (0.95) = 95$ وأن :

$$Z = \frac{97 - 95}{\sqrt{(100)(0.95)(0.05)}} = 0.917$$

(٦) الاستنتاج : بما أن قيمة $Z = 0.917$ تقع ضمن منطقة القبول ، لذلك فإننا نقبل H_0 ونستنتج أن الطرق الحديثة المستخدمة فى عملية الإنتاج لم تطور هذه العملية ولم تخفض نسبة العناصر المعابة .

(٧,٧) اختبار الفرق بين نسبتين

Testing the difference between two proportions

عندما نرغب فى اختبار فرض حول تساوى نسبتين ، فإنه تنشأ عادة حالات متعددة . فمثلا يمكننا أن نحاول البرهان على أن نسبة أطباء الأطفال فى مدينة جدة بالمملكة العربية السعودية يساوى نسبة أطباء الأطفال فى مدينة الرياض . كما يمكن لشخص أن يقرر الإقلاع عن التدخين فقط إذا كان متأكدا من أن نسبة المدخنين المصابين بسرطان الرئة يفوق نسبة غير المدخنين المصابين بنفس المرض .

نرغب بشكل عام فى اختبار فرض العدم $H_0 : p_1 = p_2 = p$ ضد فرض بديل مناسبة H_1 . حيث يمثل الوسطاء p_1 , p_2 نسبتى المجتمعين ، ويمثل المتغير العشوائى $\hat{p}_1 - \hat{p}_2$ الإحصاء الذى سنبنى عليه قرارنا . حيث يمثل $\hat{p}_1 , \hat{p}_2$ نسبتى النجاح المحسوبتين من خلال عينتين عشوائيتين مستقلتين حجماهما n_1 , n_2 على الترتيب مأخوذتين من مجتمعين حدانيين . وكما مر معنا فى (٦,٦) فإن للإحصاء :

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{(p_1q_1/n_1) + (p_2q_2/n_2)}} = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{pq\,[(1/n_1) + (1/n_2)]}}$$

توزيعاً طبيعياً معيارياً عندما يكون H_0 صحيحاً وحجما العينتين n_1 , n_2 كبيرين .

لحساب قيمة Z علينا أن نقدر المتغير p الموجود تحت إشارة الجذر . بفرض أن x_1 , x_2 يمثلان على الترتيب عدد النجاحات فى كل عينة . لذلك يمكن أن نأخذ :

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2}$$

وبتبديل $\hat{p}$ عن p فى عبارة الإحصاء Z فإننا نجد العلاقة :

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}\hat{q}\,[(1/n_1) + (1/n_2)]}}$$

حيث يمثل $\hat{q} = 1 - \hat{p}$.

لذلك وعند اختبار فرض حول تساوى نسبتى مجتمعين عندما تكون حجوم العينات المسحوبة كبيرة فإننا نتبع الخطوات الست التالية :

(١) نقوم بتحديد فرض العدم $H_0 : p_1 = p_2$

(٢) بعد ذلك نحدد الفرض البديل $p_1 \quad p_2$ أو $H_1 : p_1 > p_2 , p_1 < p_2$

(٣) نختار مستوى معنوية الاختبار α

(٤) فتكون المنطقة الحرجة هى إحدى المناطق الثلاث التالية :

أ ــ $Z < - Z_{\alpha}$ وذلك من أجل الفرض البديل $p_1 < p_2$

ب ــ $Z > Z_{\alpha}$ وذلك من أجل الفرض البديل $p_1 > p_2$

جـ ــ $Z < - Z_{\alpha/2}$ و $Z > Z_{\alpha/2}$ وذلك إذا كان الفرض البديل $p_1 \neq p_2$

(٥) ثم نحسب $\hat{p}_1 = \frac{x_1}{n_1}$, $\hat{p}_2 = \frac{x_2}{n_2}$ و $\hat{p} = \frac{x_1 + x_2}{n_1 + n_2}$ وبعدها نجد قيمة الإحصاء :

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}\hat{q}\,[(1/n_1) + (1/n_2)]}}$$

(٦) وأخيراً فإننا نتخذ قرارنا حول رفض H_0 فيما لو وقعت قيمة الإحصاء Z المحسوبة سابقا داخل المنطقة الحرجة أو أن نقبله فى غير ذلك .

مثال (٧,١٤)

يجرى التصويت على اقتراح لإقامة مصنع كيميائى بين المقيمين فى مدينة معينة والمقاطعات المحيطة بها . وسيكون المكان الذى سينشأ عليه هذا المصنع ضمن مشروع الاقتراح هو فى داخل حدود المدينة . ولهذا السبب فإن المصوتين من بين سكان المقاطعات يشعرون أن الاقتراح سيمر بسبب النسبة الكبيرة لأصوات المقترعين فى المدينة والذين يدعمون مثل هذا المشروع .

ولتحديد ما إذا كان هناك فرق هام بين نسبة أصوات المقيمين فى المدينة وأصوات المقيمين فى المقاطعات والذين يدعمون الاقتراح ، فقد جرى اقتراع جزئى ، فتبين أن 140 من أصل 250 من المقترعين فى المدينة قد صوتوا لصالح هذا الاقتراع ، وأن 290 من 600 مقترع فى المقاطعات قد صوتوا لصالحه أيضا . فهل نوافق القول بأن نسبة أصوات المقيمين فى المدينة والمؤيدين للاقتراح المذكور هى أكبر من نسبة أصوات أهل المقاطعات ؟ استخدم 0.025 مستوى معنوية .

الحل

نفرض أن p_1 , p_2 يمثلان على الترتيب نسبة المقترعين فى المدينة والمقاطعات والمصوتين لصالح الاقتراح . نلاحظ بإتباع الخطوات الست المذكورة سابقا أن :

(١) $H_0 : p_1 = p_2$

(٢) $H_1 : p_1 > p_2$

(٣) $\alpha = 0.025$

(٤) أما بالنسبة للمنطقة الحرجة فهى $Z > 1.96$

(٥) نحسب من معطيات المسألة $\hat{p}_1 = \frac{x_1}{n_1} = \frac{140}{250} = 0.56$, $\hat{p}_2 = \frac{x_2}{n_2} = \frac{290}{600} = 0.483$

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2} = 0.5058$$

لذلك فإن :

$$Z = \frac{0.56 - 0.483}{\sqrt{(0.5058)(0.4942)[1/250 + 1/600]}} = 2.045$$

(٦) نلاحظ أن قيمة الإحصاء $Z = 2.045$ تقع داخل المنطقة الحرجة ، ولذلك فإننا نرفض الفرض H_0 ونوافق القول بأن نسبة أصوات المقترعين فى المدينة لصالح الاقتراح أكبر من نسبة أصوات أهل المقاطعات المحيط بهذه المدينة .

تمارين محلولة

تمرين (١)

من المعلوم أن مقياس الذكاء فى إحدى مدارس بحرة التابعة لمدينة جدة فى المملكة العربية السعودية يتوزع وفقا للتوزيع الطبيعى بالوسط $\mu = 110$ والتباين $\sigma^2 = 100$. ولدراسة مدى اختلاف وسط الذكاء اخترنا عينة من إحدى مدارس جدة مؤلفة من 25 طالباً . فتبين أن وسط الذكاء $\bar{x} = 115$. فهل وسط الذكاء فى هذه المدرسة يختلف عن وسط الذكاء فى بحرة ؟ استخدم لذلك مستوى معنوية $\alpha = 0.1$.

الحل

لحل هذا التمرين نتبع الخطوات الست التالية :

(١) لنختبر فرض العدم $H_0 : \mu \neq 110$

(٢) ضد الفرض البديل $H_1 : \mu_x \neq 100$

(٣) وذلك تحت مستوى من المعنوية قدره $\alpha = 0.1$

(٤) نلاحظ بالعودة إلى السطر الأول فى الجدول (٧,٢) أن المنطقة الحرجة للفرضية H_0 تتحدد بالمتباينين $Z < -Z_{\alpha/2}$ و $Z > Z_{\alpha/2}$. من الجدول IV نجد أن هذه المنطقة هى Z

$Z < -1.65 , Z < 1.65$

(٥) لنحدد قيمة احصاء الاختبار :

$$Z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{115 - 110}{10 / \sqrt{25}} = 2.5$$

(٦) الاستنتاج : بما أن هذه القيمة تقع ضمن المنطقة الحرجة ، لذلك فإننا نرفض الفرض H_0 ونستنتج أن وسط الذكاء فى هذه المدرسة يختلف عنه فى منطقة بحرة .

تمرين (٢)

بفرض أن وسط الزمن اللازم للقيام بعملية صناعية معينة هو $\mu = 12.5$ دقيقة . اخترنا عشرة مستخدمين جدد ودربوا على هذه العملية ، ولدى اختبار الزمن اللازم لكل منهم للقيام بمثل هذه العملية وجدنا أن :

رقم العامل	1	2	3	4	5	6	7	8	9	10
الزمن اللازم	9.3	12.1	15.7	10.3	12.2	14.8	15.1	13.2	15.9	14.5

اختبر الفرض التالى : أن وسط الزمن اللازم لأى مستخدم لا يختلف عن وسط الزمن 12.5 دقيقة . استخدم 0.05 مستوى معنوية .

الحل

لحل هذا التمرين نتبع الخطوات الست التالية :

(١) أن فرض العدم فهو $H_0 : \mu = 12.5 \; (= \mu_0)$

(٢) أما بالنسبة للفرض البديل فهو $H_1 : \mu \neq 12.5$

(٣) أن مستوى المعنوية $\alpha = 0.05$.

(٤) لنحدد المنطقة الحرجة لإحصاء الاختبار $t = \frac{\bar{x} - \mu}{S / \sqrt{n}}$ وذلك من أجل عدد من درجات الحرية قدره $\nu = 9$. وبالعودة إلى السطر الثانى من الجدول (٧،٢) نجد أن المنطقة الحرجة للفرض H_0 هى المنطقة .

$$t < -t_{1-\alpha/2} = t_{0.975} = -2.26 \; , \; t > t_{\alpha/2} = t_{0.025} = 2.262$$

(٥) نلاحظ من معطيات المسألة أن $\bar{x} = 13.31$ وأن $S = 2.28$. لذلك فإن قيمة إحصاء الاختبار هى :

$$T = \frac{\bar{x} - \mu}{S / \sqrt{n}} = \frac{13.31 - 12.5}{(2.28) / \sqrt{10}} = 1.13$$

(٦) بما أن قيمة $T = 1.13$ لا تقع ضمن المنطقة الحرجة ، لذلك فإننا نقبل H_0 ونستنتج أنه لايوجد دليل على أن وسط الزمن اللازم للعمال الجدد للقيام بمثل هذه العملية يختلف عن الوسط المفروض (12.5 دقيقة) .

تمرين (٣)

يدعى صاحب مصنع لإنتاج الحبال أن لحباله المنتجة وسط مقاومة للقطع قدره 8000 N ، ولدراسة هذا الادعاء قمنا باختبار ستة حبال من إنتاج هذا المصنع ، فأظهرت نتائج التجربة أن لها وسط قوة مقاومة للقطع قدره 7750 N وانحرافا معياريا 145 N ، فهل يمكن تأييد ادعاء صاحب المصنع ؟ استخدم مستوى معنوية قدره 0.05 .

الحل

علينا أن نقرر بين الفرضين :

$H_0 : \mu = 8000$ N وادعاء المصنع له ما يبرره

$H_1 : \mu < 8000$ N وادعاء المصنع ليس له ما يبرره

أى أن المطلوب اختبار فرض وحيدالذيل . ونحن نجد تحت الفرض H_0 أن :

$$T = \frac{\overline{X} - \mu}{S / \sqrt{n}} = \frac{7750 - 8000}{145 / \sqrt{6}} = -4.22$$

وعند مستوى من المعنوية قدره $\alpha = 0.05$ ، أن المنطقة الحرجة للفرض H_0 حسب الجدول (٧,٢) السطر الثانى هى المنطقة $t < -t_{0.05} = -2.01$ ، وذلك من أجل $= n - 1 = 1$. وبما أن قيمة الإحصاء $t = -4.22$ تقع داخل المنطقة الحرجة ، لذلك فإننا نرفض H_0 وهذا يعين أن إدعاء صاحب المصنع ليس له ما يبرره .

تمرين (٤)

قمنا بتجربة لتقييم طريقة جديدة فى إنتاج الماس . وقد ولدنا ست ماسات بالطريقة الجديدة فكانت أوزانها 0.48 ,0.52 ,0.61 ,0.45 ,0.54 ,0.57 قيراطا . وقد بينت دراسة كلفة هذه الطريقة أنه يجب أن يكون وسط وزن الماس المستحصل أكبر من 0.5

قيراطاً لكى تكون الطريقة الجديدة مربحة . هل تقدم الأوزان السابقة دلالة كافية للقول بأن الطريقة الجديدة مقبولة ؟ استخدم مستوى معنوية $\alpha = 0.05$.

الحل

لحل هذا التمرين نقوم باتباع الخطوات الست المعروفة :

(١) أن فرض العدم هو $H_0 : \mu = 0.5$

(٢) كما أن الفرض البديل هو $H_1 : \mu > 0.5$

(٣) ومستوى معنوية الاختبار $\alpha = 0.05$

(٤) وحسب السطر الثانى فى الجدول (٧,٢) فإننا نجد أن المنطقة الحرجة هى المنطقة $t > t_{0.05}$ ، وحيث أن $= 5$. وبالعودة إلى الجدول ٧ نجد أن هذه المنطقة هى $t < 2.015$.

(٥) ونلاحظ أن قيمة إحصاء الاختبار هى :

$$t = \frac{\bar{X} - \mu_0}{S / \sqrt{n}} = \frac{0.53 - 0.5}{0.0559 / \sqrt{6}} = 1.31$$

(٦) الاستنتاج : نلاحظ أن قيمة الإحصاء المحسوب $t = 1.31$ تقع خارج منطقة الرفض ، لذلك فإننا نقبل الفرض $H_0 : \mu = 0.5$ ، وهذا يعنى أن المعلومات التى حصلنا عليها لا تقدم لنا دليلا كافيا على أن وسط وزن الماس المنتج يتجاوز 0.5 قيراطا .

تمرين (٥)

تحتاج عملية تجميع جهاز تلفزيون فى منشأة النصر لصنع التلفزيونات إلى تدريب لمدة شهر تقريبا يتلقاه مستخدم جديد للوصول إلى كفاءة عالية . وقد اقترحت طريقة جديدة للتدريب . وتم القيام بتجربة لمقارنة الطريقة الجديدة بالطريقة المعتادة . فقمنا بتدريب مجموعتين من الشباب ، تحوى كل منهما تسعة عمال جدد مستخدمين ، واحدة منهما فى كل من الطريقتين ، وذلك لمدة ثلاثة أسابيع . ثم قسنا فى النهاية الزمن اللازم بالدقائق لكل عامل لتجميع الجهاز ، فوجدنا النتائج المسجلة فى الجدول التالى :

الطريقة القديمة	الطريقة الجديدة
32	35
37	31
35	29
28	25
41	34
44	40
35	27
31	32
34	31

هل تقدم هذه النتائج معلومات كافية على تفوق الطريقة الجديدة ؟

الحل

لنفرض أن الزمن اللازم لتجميع جهاز تلفزيونى بالطريقة القديمة يتوزع وفقا للتوزيع الطبيعى بالوسط μ_1 والتباين σ^2 وبالنسبة للطريقة الجديدة بالوسط μ_2 والتباين σ^2 .

من الجدول السابق نجد أن $\bar{x}_1 = 35.22$ ، $\bar{x}_2 = 31.56$

$$\sum_{j=1}^{9} (\bar{x}_j - \bar{x}_1)^2 = 195.56 \; , \quad \sum_{i=1}^{9} (\bar{x}_i - \bar{x}_2) = 16.22$$

أما تقدير التباين المشترك فهو يساوى :

$$S^2 = \frac{\sum_{j=1}^{9} (x_j - \bar{x}_1)^2 + \sum_{i=1}^{9} (\bar{x}_i - \bar{x}_2)^2}{n_1 + n_2 - 2} = \frac{195.56 + 160.22}{9 + 9 - 2} = 22.24$$

وقيمة S هى $S = 4.71$

لنختبر فرض العدم $H_0 : \mu_1 - \mu_2 = 0$ ضد الفرض البديل $H_1 : \mu_1 - \mu_2 > 0$. من السطر الرابع فى الجدول (٧,١) وبالعودة إلى الجدول ٧ نجد أن القيمة الحرجة لـ t من أجل $\alpha = 0.05$ ، $\nu = n_1 + n_2 - 2 = 16$ درجة حرية هى القيمة $t = 1.746$ ، وتكون المنطقة الحرجة $t > 1.746$. وإذا حسبنا قيمة إحصاء الاختبار لوجدنا أن :

$$t = \frac{\bar{x}_1 - \bar{x}_2}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{35.22 - 31.56}{4.71\sqrt{\frac{1}{9} + \frac{1}{9}}} = 1.65$$

نلاحظ أن قيمة هذا الإحصاء $t = 165$ لا تقع ضمن المنطقة الحرجة ، ولذلك فإننا نقبل H_0 ونستنتج أنه لاتوجد دلالة كافية على أن الطريقة الجديدة متفوقة على القديمة .

تمرين (٦)

يريد مكتب الدراسات فى جامعة الملك عبد العزيز بجدة دراسة ما إذا كانت الحالة الاقتصادية للطالب ذات أثر فى قدراته الدراسية . ولهذه الغاية تم تقسيم الطلاب إلى فئتين وأخذت عينة عشوائية مؤلفة من مائة طالب من كل فئة وتم دراستها ، فكان وسط العينة الأولى $\bar{x} = 67.5$ وتباينها العينى $S_1^2 = 225$. كما وجد أن $\bar{x}_2 = 63.5$ وأن $S_2^2 = 250$. هل تقدم هذه المعلومات دلالة كافية على وجود فرق فى القدرات الدراسية بين الفئتين ؟ استخدم لذلك مستوى معنوية $\alpha = 0.05$.

الحل

لنرمز للمعدل العام لطالب الفئة الأولى بالرمز μ_1 ، ولطالب الفئة الثانية بالرمز μ_2 . لنتبع الخطوات الستة المعروفة :

(١) أن فرض العدم هو $H_0 : \mu_1 - \mu_2 = d_0, (d_0 = 0)$

(٢) أما الفرض البديل فهو $H_1 : \mu_1 - \mu_2 \neq d_0$

(٣) بالنسبة لمستوى معنوية الاختبار فهو يساوى $\alpha = 0.05$

(٤) نعلم أن الإحصاء $\bar{x}_1 - \bar{x}_2$ يمثل تقديراً جيداً للفرق $\mu_1 - \mu_2$ وله توزيع قريب من التوزيع الطبيعى (لأن $n_1 = 100, n_2 = 100$) بوسط قدره $\mu_1 - \mu_2$ وتباين قدره :

$$\sigma^2_{\bar{x}_1 - \bar{x}_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

بالعودة إلى السطر الثالث فى الجدول (٧,٢) نجد أن المنطقة الحرجة للإحصاء :

$Z = \frac{(\bar{X}_1 - \bar{X}_2) - d_0}{\sqrt{(\sigma_1^2/n_1) + (\sigma_1^2/n)}}$ هى المنطقة $Z > 1.96, Z < -1.96$.

(٥) لنحسب قيمة إحصاء الاختبار فنجد أن :

$$Z = \frac{67.5 - 63.5}{\sqrt{(225/100) + (250/100)}} = 1.83$$

(٦) بما أن قيمة هذا الإحصاء لا تقع ضمن المنطقة الحرجة ، لذلك فإننا نقبل H_0 ونستنتج أن الحالة الاقتصادية لا علاقة لها بقدرات الطالب الدراسية تحت مستوى معنوية قدره $\alpha = 0.05$.

تمرين (٧)

تبين سجلات مستشفى جامعة الملك عبد العزيز بجدة أن 52 رجلا من عينة مكونة من 1000 رجل يقابلها 23 إمرأة من أصل 1000 إمرأة ممن كانوا يعانون من مرض فى المعدة خلال العام ١٤٠٠ — ١٤٠١ هجرية . هل تقدم هذه المعلومات الإحصائية دلالة كافية على أن نسبة المرضى بين الرجال بأمراض المعدة هو أكبر منها عند النساء ؟

الحل

نفرض أن عدد المرضى الوافدين إلى قسم الأمراض المعدية فى المستشفى خلال عام سواء من الرجال أو النساء يتبع التوزيع الحدانى بوسطين هما على الترتيب p_2, p_2 لنختبر فرض العدم $H_0 : P_1 = P_2$ ضد الفرض البديل $H_0 : P_1 \quad P_2$ وكتقدير للفرض $P_1 - P_2$ نأخذ التقدير النقطي $\hat{p}_1 - \hat{p}_2$. نعلم أن لهذا التقدير توزيعاً طبيعيا بالوسط $p_1 - p_2$ والتباين $\sigma^2_{(\hat{p}_1 - \hat{p}_2)} = \frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}$ وكإحصاء لاختبار الفرض H_0 نأخذ المتغير العشوائي

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sigma(\hat{p}_1 - \hat{p}_2)}$$

فإذا استخدمنا قيم تقريبية مناسبة لكل من p_2 , p_1 فى عبارة $\sigma^2_{(\hat{p}_1 - \hat{p}_2)}$ فإننا نجد أن التقدير المشترك هو :

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{52 + 23}{1000 + 1000} = 0.0375$$

وإحصاء الاختبار هو :

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}\hat{q}\,(\frac{1}{n_1} + \frac{1}{n_2})}} = \frac{0.52 - 0.023}{\sqrt{(0.0375)\,(0.9625)\,(1/1000 + 1/1000)}}$$

وتكون منطقة الرفض من أجل $\alpha = 0.05$ هى المنطقة $Z > 1.96$ و $Z < -1.96$. نلاحظ أن قيمة الإحصاء السابقة تقع ضمن هذه المنطقة الحرجة لذلك نرفض الفرض $p_1 = p_2$ ونستنتج أن المعلومات الإحصائية السابقة تقدم لنا دلالة قوية على أن نسبة المرض بأمراض المعدة من الرجال كانت أكبر منها عند النساء ممن راجعوا المستشفى المذكور خلال نفس العام .

تمرين (٨)

جرى اختبار لدراسة تأثير سماد من نوع معين على إنتاج الشعير . ولهذا الغرض اختيرت 24 قطعة من الأرض لها نفس المساحة ، وعولج نصفها بالسماد بينما ترك النصف الآخر بدون معالجة . ولدى جنى المحصول وجد أن وسط الغلة من الشعير فى مجموعة القطع المتروكة هو 4.8 طناً ، وأن انحرافها المعيارى هو 4 طن ، بينما كان وسط غلة الفدان للقطع التى تمت معالجتها بالسماد هو 5.1 طنا وانحرافها المعيارى 3.6 طنا . هل هناك تحسن معنوى فى إنتاج الشعير نتيجة استخدام السماد ، وذلك باستخدام مستوى معنوية قدره $\alpha = 0.01$ ؟

الحل

لنفرض أن μ_1 , μ_2 تمثل وسط مجتمع غلة الشعير من الأرض المعالجة والأرض غير المعالجة على الترتيب ، علينا أن نقرر بين الفرضين :

$$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 > \mu_2$$

والسماد يؤدى إلى تحسين الغلة وتحت فرض العدم H_0 نأخذ الإحصاء

$T = \dfrac{\bar{X}_1 - \bar{X}_2}{S_p \cdot \sqrt{1/n_1 + 1/n_2}}$ من أجل $22 = n_1 + n_2 - 2 =$ درجة من الحرية ، وحيث تمثل

$$S_p^2 = \frac{(n_1 - 1)\, S_1^2 + (n_2 - 1)\, S_2^2}{n_1 + n_2 - 2}$$

وهكذا نجد أن :

$$S_p^2 = \frac{(11)\,(4)^2 + (11)\,(3.6)^2}{12 + 12 - 2} = 14.48$$

أما قيمة إحصاء الاختبار T فهى تساوى :

$$T = \frac{5.1 - 4.8}{(3.80)\sqrt{1/12 + 1/12}} = 0.193$$

من الجدول (٧,٢) السطر الرابع نجد أن المنطقة الحرجة للفرض H_0 هى $T > t_a =$ $t_{0.01}$ بـ $22 =$ درجة من الحرية ، وهذه المنطقة هى $T > 2.508$. وبما أن قيمة الإحصاء
$T = 0.193$ لا تقع ضمن المنطقة الحرجة لذلك فإننا نقبل H_0 .

تمرين (٩)

كان الانحراف المعيارى فى فترات سابقة لأوزان عبوات تزن 40.0 N تملأ بواسطة آلة معينة هو 0.25N ولدراسة الزيادة الظاهرة فى التباين اخترنا عينة عشوائية مؤلفة من 20 عبوة ، فكان انحرافها المعيارى 0.32N . فهل هذا يدل على زيادة ظاهرة فى التباين ؟ استخدم مستوى معنوية قدره 0.05 .

الحل

لحل هذا التمرين نتبع الخطوات الست المعروفة :

(١) أن فرض العدم هو $H_0 : \sigma = 0.25$

(٢) أما الفرض البديل فهو $H_1 : \sigma > 0.25$ وهذا يعنى أن هناك زيادة فى التباين .
(٣) وبالنسبة لمستوى معنوية الاختبار فهو $\alpha = 0.05$
(٤) وحسب الجدول (٧,٢) السطر السادس نجد أن المنطقة الحرجة فى هذا الاختبار هى المنطقة المحددة بالمتباينة $\chi^2 > \chi^2_\alpha$ علماً بأن عدد درجات الحرية لتوزيع كاى – مربع هو $\nu = n - 1 = 19$ وعلى هذا فإن المنطقة الحرجة هى $\chi^2 > 30.144$ وذلك بالعودة إلى الجدول VI .
(٥) لنحسب قيمة إحصاء الاختبار :

$$\chi^2 = \frac{(n - 1) S^2}{\sigma_0^2} = \frac{(19)(0.32)^2}{(0.25)^2} = 31.2$$

(٦) الاستنتاج : بما أن قيمة إحصاء الاختبار تقع داخل المنطقة الحرجة ، لذلك فإننا نرفض H_0 بمستوى معنوية $\alpha = 0.05$ ونقبل H_1 ، وهذا يعنى أن هناك زيادة ظاهرة فى التباين .

تمرين (١٠)

عينتان عشوائيتان مستقلتان حجماهما $n_2 = 10, n_1 = 8$ مأخوذتان من مجتمعين تباينهما σ_2^2 , σ_1^2 على الترتيب . حسبنا من خلال هاتين العينتين كلا من S_1^2 , S_2^2 فوجدنا أن $S_1^2 = 7.14, S_2^2 = 3.21$. هل تقدم هذه القيم دلالة كافية على عدم تساوى تباينى المجتمعين المذكورين ؟ افرض أن توزع المجتمعين المدروسين طبيعيين .

الحل

لحل هذا التمرين نتبع الخطوات الست المعروفة لدراسة اختبار ما

(١) نفرض أن $H_0 : \sigma_1^2 = \sigma_2^2$
(٢) وأن الفرض البديل هو $H_1 : \sigma_1^2 \neq \sigma_2^2$
(٣) لنفرض أن مستوى معنوية الاختبار هو $\alpha = 0.1$
(٤) بالعودة إلى الجدول (٧,٢) السطر الأخير نجد أن المنطقة $F < f_{1-\alpha/2}(\nu_1, \nu_2)$ و $F > f_{\alpha/2}(\nu_1, \nu_2)$ تحدد لنا ما يسمى بالمنطقة الحرجة حيث رمزنا لإحصاء الاختبار بالرمز $F = \frac{S_1^2}{S_2^2}$ حيث يمثل $\nu_1 = n_1 - 1 = 7$ و $\nu_2 = n_2 - 1 = 9$ ، وبالعودة إلى الجدول VII نجد أن

$$f_{\alpha/2}(\nu_1 , \nu_2) = f_{0.05}(7, 9) = 3.29$$

وبما أن $f_{1-\alpha}(\nu_1 , \nu_2) = \dfrac{1}{f_\alpha(\nu_2 , \nu_1)}$ فمن الجدول VII نجد أيضا أن :

$$f_{1-\alpha/2}(\nu_1 , \nu_2) = f_{0.95}(7 , 9) = \frac{1}{f_{0.05}(9 , 7)} = \frac{1}{3.68} = 0.27$$

والمنطقة الحرجة فى هذا الاختبار هى $F < 0.27$, $F > 3.29$ نلاحظ أن قيمة إحصاء الاختبار :

$$F = \frac{S_1^2}{S_2^2} = \frac{7.14}{3.21} = 2.22$$

وحيث أن هذه القيمة لا تقع ضمن منطقة الرفض ، لذلك نقبل فرض العدم ، وهذا يعنى أنه لايوجد دلالة كافية على وجود فرق بين التباينين .



الفصل الثامن

الانحدار والارتباط

● تمهيد ■ الانحدار الخطى ■ الانحدار الخطى البسيط ■ خواص تقديرات المربعات الصغرى ■ نهايات الثقة واختبارات المعنوية ■ تحليل التباين ■ القياسات المتكررة لـ y ■ الارتباط.



(٨,١) تمهيد

سندرس فى هذا الفصل السلوك المشترك لمتغيرين أو أكثر ، فغالباً ما يهتم باحث ما أو مجرب بدراسة ما إذا كانت هناك علاقة بين متغيرين أو أكثر ، ومن ثم يلجأ إلى التعبير عن هذه العلاقة بصورة رياضية وذلك بتحديد المعادلة التى تربط هذه المتغيرات .

فمثلا يدرس الفيزيائى العلاقة بين ضغط وزن معين من الغاز (والذى يعتمد على درجة حرارته) وحجم الغاز ، وكذلك يدرس المهندس الزراعى العلاقة بين كمية محصول غلة القمح وكمية السماد التى يتم بها تزويد نبات القمح فى نفس العام ، كما تهتم إدارات الجامعات بمسألة الإنجاز المتوقع لطالب عند نهاية سنته الجامعية الأولى ، فقد ترغب جامعة ما فى تقدير ما سيكون عليه معدل كل طالب فى سنته الجامعية الأولى وذلك قبل التسجيل . والملاحظ أن معدل الطالب سيكون تابعا لعدة متغيرات . مثلا الدرجة التى حصل عليها هذا الطالب خلال امتحان الثانوية فى مادة معينة ، مرتبة نجاحه فى الثانوية ، وكذلك درجة الاختبار التمهيدى الذى تجريه الجامعة فى نفس المادة .

نلاحظ أن جميع المسائل المطروحة سابقا تمثل مسائل طبيعية عامة جداً ، فنحن نهتم بالعلاقة بين متغير عشوائى y – مثلا – وعدد من المتغيرات المستقلة وغير العشوائية فى طبيعتها ، مثل $x_1, x_2, \ldots, x_n$. ففى المثال الأخير كان المتغير العشوائى y ممثلا لمعدل الطالب فى امتحان الرياضيات مثلا ، أما المتغيرات x_1, x_2, x_3 فهى على الترتيب درجة الرياضيات فى امتحان الشهادة الثانوية ، مرتبة الطالب ، ودرجة الطالب فى الامتحان التمهيدى . إن هدف الجامعة هو قياس x_1, x_2, x_3 من أجل طالب معين ، ثم تبديل هذه القياسات فى معادلة تنبؤ تحاول الحصول عليها لتحصل بذلك على تنبؤ عما سيكون عليه معدل الطالب فى نهاية السنة الجامعية الأولى . إذا لا بد قبل كل شىء من تحديد المتغيرات المستقلة $x_1, \ldots, x_n$. ثم الحصول على قياس لمدى علاقتها بالمتغير y ، وبعد ذلك وضع معادلة تنبؤ تعبر عن y كدالة بهذه المتغيرات المستقلة . تُعرف مسألة تنبؤ متغير بمعرفة متغيرات أخرى بمسألة الانحدار .

(٨,٢) الانحدار الخطى The linear regression

تعريف (٨,١) البيانات الثنائية

بفرض أن $x_1, x_2, ..., x_n$ تمثل قيم متغير x ، $y_1, ..., y_n$ القيم المقابلة لمتغير ثان y ، عندئذ تسمى مجموعة الأزواج : (x_1, y_1) , (x_2, y_2) ,...., (x_n, y_n) بيانات ثنائية .

مثال (٨,١)

لنفرض أن القيم $x_1, x_2, ..., x_{25}$ تمثل أوزان 25 شخصا فى مدينة جدة . ولنفرض أن $y_1, y_2, ..., y_{25}$ تمثل أوزان أكبر أولاد هؤلاء الأشخاص على الترتيب ، عندئذ تمثل مجموعة الأزواج (x_1, y_1) ... (x_{25}, y_{25}) بيانات ثنائية .

مثال (٨,٢)

بفرض أن القيم $x_1, ..., x_{10}$ تمثل الدرجة التى نالها كل طالب (من بين كل عشرة طلاب) فى مادة الرياضيات فى امتحان الشهاد الثانوية ، وأن $y_1, y_2, ..., y_{10}$ تمثل القيم الموافقة التى حصل عليها كل طالب من هؤلاء الطلاب خلال الامتحان التمهيدى الذى أجرته الجامعة فى هذه المادة عند فحص القبول . عندئذ تمثل الأزواج (x_1, y_1) ,...., (x_{10}, y_{10}) مجموعة بيانات ثنائية .

لدراسة العلاقة الممكنة مثلا بين درجة الرياضيات x فى امتحان الثانوية ، ودرجة الرياضيات فى الفحص التمهيدى فى الجامعة ، فإننا نقوم بدراسة العلاقة الممكنة بين أزواج قيم x وقيم y ، وذلك بنشر هذه الأزواج فى أماكنها المحددة فى المستوى الاحداثى . وعندئذ يمثل المخطط الناتج عن هذه النقاط ما يسمى بمخطط الانتشار (scatter diagram) .

إن إحدى الطرق الممكنة للحصول على معادلة تنبؤ تربط بين y و x هو وضع مسطرة فوق التمثيل البيانى ثم تحريكها حتى تبدو وكأنها تمر عبر أكبر عدد ممكن من النقاط ، وهذا الوضع يقدم لنا ما يسمى بأفضل تلاؤم (best fit) مع المعلومات

الإحصائية المتوافرة . وعند رسم هذا المستقيم عبر هذه النقاط تكون مسألة التنبؤ المطروحة منتهية ، حيث إنه بإمكاننا استخدام هذا الخط للتنبؤ بدرجة الطالب y فى الرياضيات عند انتهاء السنة الأولى ، وذلك من أجل قيمة معينة لدرجته x فى اختبار الثانوية .

نلاحظ فى المثال السابق أننا اخترنا نموذجا رياضيا يعبر عن العلاقة المفروضة بين x و y ، ومن المعروف أنه بإمكاننا التعبير عن معادلة أى مستقيم على الشكل :

$$y = a + bx$$

حيث يمثل a ترتيب نقطة تقاطع هذا المستقيم مع المحور الرأسى ، كما يمثل b ميل هذا المستقيم ، وكما نعلم فإنه يقابل كل مستقيم معادلة خطية بسيطة من هذا النوع والعكس صحيح . والملاحظ أنه عندما نقوم برسم مستقيم عبر الأزواج نكون قد اخترنا وبشكل آلى معادلة رياضية .

$$y = a + bx$$

وبهذا الشكل تتحد قيمتا a و b بصورة وحيدة . نسمى النموذج الخطى $y = a + bx$ بنموذج رياضى حتمى لأنه عندما نبدل قيمة x فى المعادلة نحصل على قيمة محددة لـ y دون أن يكون هناك أى مجال للخطأ .

إن النماذج الحتمية ملائمة لشرح كثير من الظواهر الفيزيائية والتنبؤ بها عندما يكون خطأ هذا التنبؤ مهملا من وجهة النظر العملية . فمثلا يسمح لنا قانون نيوتن $F = m.a$ (حيث تمثل F القوة التى ينتجها جسم متحرك كتلته m ، وتسارعه a) فى كثير من الحالات العملية بالتنبؤ بالقوة F بخطأ صغير مهمل عملياً ، واعتبار الخطأ صغيراً أو كبيراً هو مسألة نسبية . فقد يكون خطأ قدره نيوتن واحداً صغيراً جدا عند إنشاء أساس لجسر ، إلا أنه خطأ كبير جدا بالنسبة لصنع قاعدة إطلاق صاروخ متجه إلى القمر . ولا يمكننا تجاهل هذا الخطأ فى عدد كبير من التجارب الفيزيائية ، وبهذا فإننا سنتردد كثيراً جدا فى منح الكثير من الثقة لتنبؤ غير مصحوب بقياس لجودة هذا التنبؤ . ولهذا فإن طريقة المسطرة والنظر لاختبار خط مستقيم يربط بين قيم x و y هى طريقة غير مقبولة ، ومحدودة الفائدة .

لتفسير بعض الظواهر الفيزيائية يمكننا استخدام النماذج الرياضية الاحتمالية عوضا عن النماذج الحتمية ، ومن المعلوم أن النموذج الرياضي الاحتمالي يحوى متغيراً أو أكثر من المتغيرات ذات الطبيعة العشوائية والتى لها توزيعات احتمالية محددة .

من الملائم تعريف المتغير العشوائى Y الموافق لقيمة ثابتة لـ x والذى نرمز له بالرمز $Y|x$ ، ولتوزيعه الاحتمالى بالرمز $f(Y|x)$. من الواضح أنه إذا كان $x = x_i$ فعندئذ يمثل الرمز $Y|x_i$ المتغير العشوائى Y_i . أن تعبير الانحدار الخطى يدل حتما على أن وسط $Y|x$ ارتباطا خطيا بـ x بواسطة علاقة ميل التقاطع العادية .

$$\mu_{Y|x} = \lambda = \beta x$$

حيث يمثل كلاً من λ و β وسيطين يطلب تقديرهما من بيانات العينة . فإذا رمزنا لتقديرى هذين الوسيطين بالرمزين a و b على الترتيب ، فعندئذ يكون تقدير المتغير المرتبط y والذى نرمز له بالرمز $\hat{y}$ محددا بعلاقة خط الانحدار العينى .

$$\hat{y} = a + bx$$

(٨,٣) الانحدار الخطي البسيط The simple linear regression

سنحاول فى هذه الفقرة تطوير إجراءات إيجاد تقدير ثوابت الانحدار λ و β ، وذلك من أجل جعل معادلة الانحدار قابلة للاستخدام فى معادلة التنبؤ أو فى معادلة تقدير وسط Y من أجل قيمة معينة للمتغير المستقل x .

فإذا فرضنا أن تجربتنا تضم متغيراً واحداً x وآخر Y ، عندئذ تأخذ البيانات شكل الأزواج : $\{(x_i , y_i)\ i = 1, 2, ..., n\}$. فإذا خضعت قيم x للمراقبة (أى للفحص) أى أن التجربة قد صممت عندئذ تكون العملية التجريبية هى اختيار القيم x_i سلفا ، ومن ثم مراقبة القيم الموافقة Y_i .

لنفرض أن كل الأوساط $\mu_{Y|x}$ تقع على مستقيم (أى أن هذه الأوساط ترتبط ارتباطاً خطياً بـ x) ولنفرض أن القيم الملاحظة (المراقبة) لـ Y ستحيد عن هذا الخط المستقيم والذى يمثل هذه العلاقة الخطية (أى ستقع تحت أو فوق الخط المستقيم) بمقدار عشوائى E يسمى بالخطأ ، عندئذ يمكن كتابة هذا المتغير $Y_i = Y|x_i$ على النحو التالى :

$$Y_i = \mu_{Y|x_i} + E_i = \lambda + \beta x_i + E_i$$

حيث يمثل E_i متغيراً عشوائياً وسطه حتما معدوم . وتحقق كل ملاحظة (x_i , y_i) فى العينة العلاقة :

$$y_i = \lambda + \beta x_i + \varepsilon_i$$

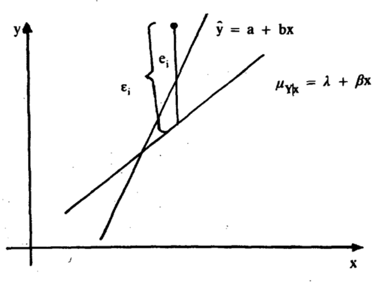
حيث يمثل ε_i قيمة يفترضها المتغير E_i ، وذلك عندما يفترض Y_i القيمة y_i ، وبصورة مماثلة وباستخدام خط تقدير الانحدار يمكننا كتابة :

$$\hat{y} = a + bx$$

وتحقق كل ملاحظة من الملاحظات (x_i , y_i) العلاقة :

$$y_i = a + bx_i + e_i$$

حيث يسمى e_i بالباقى . يوضح الشكل (٨,١) الفرق بين ε_i , e_i



الشكل (٨,١)

سنحاول إيجاد b , a ، وهما تقديرى λ , β بحيث يكون مجموع مربعات البواقى أصغريا ، نسمى عادة مجموعات مربعات البواقى بمجموع مربعات الأخطاء حول خط الانحدار ونرمز له بالرمز SSE . كما نسمى هذه الطريقة بطريقة المربعات الصغرى ، وبما أننا سنحاول إيجاد كل من b , a بحيث يكون المقدار التالى أصغرياً :

$$SSE = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - a - bx_i)^2$$

لذلك فإننا سنقوم أولا باشتقاق SSE جزئيا بالنسبة لـ a ، ثم بالنسبة لـ b حيث نجد أن :

$$\frac{\partial(SSE)}{\partial a} = -2 \sum_{i=1}^{n} (y_i - a - bx_i)$$

$$\frac{\partial(SSE)}{\partial a} = -2 \sum_{i=1}^{n} (y_i - a - bx_i)x_i$$

وبمساواة كل مشتقة من هاتين المشتقتين الجزئيتين بالصفر ، فإننا نحصل على المعادلتين :

$$na + b \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$$

$$a \sum_{i=1}^{n} x_i + b \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i$$

اللتين تسميان بالمعادلتين الطبيعيتين واللتين تقودان إلى :

$$b = \frac{n \sum_{i=1}^{n} x_i y_i - (\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{n \sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2}$$

ومن المعادلة الطبيعية الأولى نجد أن :

$$a = \bar{y} - b\bar{x}$$

تعريف (٨,٢) تقديرى الوسيطين β , λ **Estimating the medians β , λ**

إن تقدير خط الانحــدار $\mu_{Y|x} = \lambda + \beta x$ والمحســوب من خلال العينة $\{(x_i , y_i) : i = 1,2, \ldots, n\}$ هو المستقيم :

$$\hat{y} = a + bx$$

حيث يمثل :

$$b = \frac{n \sum_{i=1}^{n} x_i y_i - (\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{n \sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2}$$

$$a = \bar{y} - b\bar{x}$$

مثال (٨,٣)

لدى مهندس كيميائى عشر قراءات لدرجة الحرارة ، فإذا علمت أن الفحوص التجريبية لدرجة الحرارة قد أعطت :

جدول (٨,١)

درجة الحرارة F°	x	100	110	120	130	140	150	160	170	180	190
درجة الحرارة التى أوضحتها التجربة	y	45	51	54	61	66	70	74	78	85	89

ما هى العلاقة الخطية بين هذه البيانات (أى ما هو انحدار y على x) ؟

الحل

نلاحظ من شروط المسألة أن :

$$\sum_{i=1}^{n} x_i = 1450 , \sum_{i=1}^{10} y_i = 673 , \bar{y} = 67.3$$

$$\sum_{i=i}^{n} x_1^2 = 218500 \quad \sum_{i=1}^{n} x_i y_i = 101570 , \bar{x} = 145$$

لذلك فإن :

$$b = \frac{10\,(101570) - (1450)\,(673)}{10\,(218500) - (1450)^2} = 0.483$$

$$a = 67.3 - (0.483)\,(145) = -2.735$$

وهكذا فإن الخط المطلوب هو :

$$\hat{y} = -\,2.732 + 0.483\,x$$

وهو يمثل خط انحدار y على x .

مثال (٨,٤)

أوجد تقديراً لخط انحدار x على y للمعلومات التالية :

x	1.5	1.8	2.4	3.0	3.5	3.9	4.4	4.8	5.0
y	4.8	5.7	7.0	8.3	10.9	12.4	13.1	13,6	15.3

الحل

نلاحظ أن :

$$\sum_{i=1}^{9} x_i = 30.3 , \sum_{i=1}^{9} x_i^2 = 115.11 , \bar{x} = 3.3667$$

$$\sum_{i=1}^{9} y_i = 91.1 \ , \ \sum_{i=1}^{9} x_i y_i = 345.09 \ , \ \bar{y} = 10.1222$$

وبالتعويض فى التعريف (٨,٢) نجد أن :

$$b = \frac{(9)(345.09) - (30.3)(91.1)}{(9)(115.11) - (30.3)^2} = 2.9303$$

$$a = 10.1222 - (2.9303)(3.3667) = 0.2568$$

وهكذا فإن تقدير خط انحدار y على x هو :

$$\hat{y} = 0.2568 + 2.9303\,x$$

(٨,٤) خواص تقديرات المربعات الصغرى

Properties of the least squares estimators

لقد فرضنا سابقا أن الخطأ فى العلاقة :

$$y_i = \lambda + \beta x_i + E_i$$

يمثل متغيراً توقعه الرياضى صفراً ، لنفرض إضافة لذلك أن لكل متغير من المتغيرات E_i توزيعاً طبيعياً بالتباين σ^2 ، وأن هذه المتغيرات $E_1, E_2, \ldots, E_n$ مستقلة فيما بينها ، لنفتش عن وسط وتباين تقديرى λ , β .

يجب ألا يغيب عن أذهاننا بأن قيم b, a الممثلان لتقديرى β, λ على الترتيب تحسب من خلال عينة n من الملاحظات ، وأنه يمكن النظر إلى التقديرات المختلفة لـ β, λ (المحسوبة من عينات مختلفة لها نفس الحجم n) كقيم يفترضها المتغيران العشوائيان A و B . وحيث أن قيم x تظل ثابتة ، لذا فإن قيم A و B تتعلق باختلافات قيم y أو بشكل أدق بقيم المتغيرات العشوائية $Y_1, Y_2, \ldots, Y_n$. إن الفرضيات التوزيعية لـ E_i تؤكد على

أن Y_i (حيث $i = 1,2 ,..., n$) تتوزع بطريقة مماثلة ، وأن لكل منها توزيعاً احتمالياً طبيعيّاً بالوسط $\lambda + \beta x_i$ والتباين σ^2 وحيث إن التقدير :

$$B = \frac{n \sum_{i=1}^{n} x_i y_i - (\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{n \sum_{i=1}^{n} x_i^2 (\sum_{i=1}^{n} x_i)^2}$$

$$= \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{Y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

$$= \frac{\sum_{i=1}^{n} (x_i - \bar{x}) Y_i}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

يمثل دالة خطية بالمتغيرات العشوائية $Y_1 , Y_2 ,..., Y_n$ والأمثال :

$$a_i = \frac{x_i - \bar{x}}{\sum_{i=1}^{n} (x_i - \bar{x})^2} , i = 1, 2, ... , n$$

فإنا نستنتج من النظرية (٥,١١) أن للمتغير B توزيعاً طبيعياً بالوسط :

$$\mu_B = E(B) = \frac{\sum_{i=1}^{n} (x_i - \bar{x}) E(Y_i)}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

$$= \frac{\sum_{i=1}^{n} (x_i - \bar{x})(\lambda + \beta x_i)}{\sum_{i=1}^{n} (x_i - \bar{x})^2} = \beta$$

والتباين :

$$\sigma_B^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2 \sigma_{Y_i}^2}{\left[\sum_{i=1}^{n} (x_i - \bar{x})^2 \right]^2} = \frac{\sigma^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

يمكننا أن نبين أن المتغير العشوائى A هو أيضا متغير عشوائى وسطه :

$$\mu_A = \lambda$$

وتباينه :

$$\sigma_A^2 = \frac{\sum_{i=1}^{n} x_i^2}{n \sum_{i=1}^{n} (x_i - \bar{x})^2}$$

ولكى نتمكن من بناء استدلال حول λ و β فإن من الضرورى الوصول إلى تقدير للتباين σ^2 الذى يظهر فى عبارتى تباين A و B السابقتين ، ومن المناسب لإيجاد مثل هذا التقدير تعريف الرموز التالية :

$$S_{xx} = \sum_{i=1}^{n} (x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - \frac{(\sum_{i=1}^{n} x_i)^2}{n}$$

$$S_{yy} = \sum_{i=1}^{n} (y_i - \bar{y})^2 = \sum_{i=1}^{n} y_i^2 - \frac{(\sum_{i=1}^{n} y_i)^2}{n}$$

$$S_{xy} = \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^{n} x_i y_i - \frac{(\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{n}$$

وهكذا نستطيع كتابة مجموع مربعات الأخطاء على النحو التالى :

$$SSE = \sum_{i=1}^{n} (y_i - a - bx_i)^2$$

$$= \sum_{i=1}^{n} [(y_i - \bar{y}) - b(x_i - \bar{x})]^2$$

$$= \sum_{i=1}^{n} (y_i - \bar{y})^2 - 2b\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) + b^2 \sum_{i=1}^{n} (x_i - \bar{x})^2$$

$$= S_{yy} - 2b\, S_{xy} + b^2 S_{xx}$$

$$= S_{yy} - b\, S_{xy} ,$$

حيث : $b = \dfrac{S_{xy}}{S_{xx}}$

يمكننا بعد هذا أن نسوق ونبرهن النظرية التالية :

نظرية

إن التقدير غير المتحيز لـ σ^2 يعطى بالعلاقة :

$$S^2 = \frac{SSE}{n-2} = \frac{S_{yy} - b\, S_{xy}}{n-2}$$

البرهان

بتفسير مجموع مربعات الأخطاء كمتغير عشوائي تتغير قيمة بإعادة التجربة مرات عديدة ، عندئذ بإمكاننا كتابة :

$$SSE = S_{yy} - B\, S_{xy}$$

$$= S_{yy} - B^2 S_{xx}$$

لأن : $(S_{xy} = B\, S_{xx})$

$$= \sum_{i=1}^{n} (y_i - \bar{y})^2 - B^2 \sum_{i=1}^{n} (x_i - \bar{x})^2$$

$$= \sum_{i=1}^{n} y_i^2 - n\,\overline{Y}^2 - \left(\sum_{i=1}^{n} x_i^2 - n\,\bar{x}^2\right) E(B^2)$$

وبأخذ التوقع الرياضى للطرفين نجد أن :

$$E\,(SSE) = \sum_{i=1}^{n} E y_i^2 - n\,E(\overline{Y}^2)$$

$$- \left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right) E(B^2)$$

وحيث إن :

$$E\ y_i^2 = \sigma_{Y_i}^2 + \mu_{Y_i}^2$$

$$E\,\overline{Y}^2 = \sigma_{\overline{Y}}^2 + \mu_{\overline{Y}}^2$$

$$E\,(B^2) = \sigma_B^2 + \mu_B^2$$

وبالتبديل فى المعادلة السابقة فإننا نجد أن :

$$E\,(SSE) = \sum_{i=1}^{n} (\sigma^2_{Y_i} + \mu^2_{Y_i}) - n\,\sigma^2_{\overline{Y}} + \mu^2_{\overline{Y}}$$

$$- \left(\sum_{i=1}^{n} x_i^2 - n\,\bar{x}^2\right)(\sigma^2_B + \mu^2_B)$$

$$= n\sigma^2 + \sum_{i=1}^{n} (\lambda + \beta x_i)^2 - n\left[\frac{\sigma^2}{n} + (\lambda + \beta\overline{X})^2\right]$$

$$- \left(\sum_{i=1}^{n} x_i^2 - n\,\bar{x}^2\right)\left(\frac{\sigma^2}{S_{xx}} + \beta^2\right)$$

$$= (n - 2)\,\sigma^2$$

لذلك فإن :

$$E(S^2) = \frac{E(SSE)}{n-2} = \sigma^2$$

أى إن S^2 يمثل تقديرا غير متحيز لـ σ^2 .

(٨,٥) نهايات الثقة واختبارات المعنوية

Confidence limits and tests of significance

لابد للدارس قبل كل شيء من أن يستقرىء ما إذا كانت هناك أصلا علاقة بين Y و x . وبتعبير آخر هل تقدم المعلومات الإحصائية المتوفرة لديه دليلا كافيا على وجود علاقة خطية بين Y و x فوق فترة معينة لقيم x ؟ والسؤال المطروح يتعلق بقيمة β .

إن قولنا بأن Y و x لا يرتبطان ببعضهما خطيا يكافىء القول بأن $\beta = 0$ ، وهكذا نخلص إلى القول فى أننا نرغب فى اختبار الفرضية $\beta = 0$ ضد الفرضية البديلة $\beta \neq 0$ ، وحيث إن β يمثل متغيراً عشوائياً طبيعياً وأن $\frac{(n-2)S^2}{\sigma^2}$ يمثل متغيراً فى نوع كأى – مربع بـ $(n-2)$ درجة من الحرية لذلك وحسب النظرية (٥,٧) فإن المتغير :

$$T = \frac{(B-\beta)/(\sigma/\sqrt{S_{xx}})}{S/\sigma} = \frac{B-\beta}{S/\sqrt{S_{xx}}}$$

يتوزع وفقا للتوزيع t بـ $(n-2)$ درجة من الحرية يمكننا استخدام الإحصاء T لايجاد $100\%(1-\alpha)$ فترة ثقة للمعامل β فى معادلة خط الانحدار $\mu_{Y|x} = \lambda + \beta x$ هى الفترة :

$$b - \frac{t_{\alpha/2} \cdot S}{\sqrt{S_{xx}}} < \beta < b + \frac{t_{\alpha/2} S}{\sqrt{S_{xx}}}$$

حيث تمثل $t_{a/2}$ قيمة الإحصاء t بـ $(n-2)$ درجة من الحرية .

مثال (٨,٥)

أجريت تجربة صندوق القص لتحديد العلاقة بين الإجهاد المتعامد (normal stress) وصامد القص (Shear resistance) للتربة فأعطت النتائج التالية :

x	الإجهاد المتعامد (KN/m²)	11	13	15	17	19	21
y	صامد القص (KN/m²)	15.2	17.7	19.3	21.5	23.9	25.4

أوجد فى معادلة الانحدار $\mu_{Y|x} = \lambda + \beta x$ 95% فترة ثقة للمعامل β .

الحل

نلاحظ أن :

$$\sum_{i=1}^{6} x_i = 96 , \sum_{i=1}^{6} y_i = 123 , \sum_{i=1}^{6} y_i^2 = 2595.44$$

$$\sum_{i=1}^{6} x_i^2 = 1606 , \sum_{i=1}^{6} x_i y_i = 2039.8$$

لذلك فإن :

$$S_{xx} = 1606 - \frac{(96)^2}{6} = 70$$

$$S_{yy} = 2595.44 - \frac{(123)^2}{6} = 73.94$$

$$S_{xy} = 2039.8 - \frac{(96)(123)}{6} = 71.8$$

وهذا يعنى بأن $b = 1.0257$ وهكذا فإن :

$$S^2 = \frac{S_{yy} - b\,S_{xy}}{n-2}$$

$$= \frac{73.94 - (1.0257)(71.8)}{4} = 0.073685$$

وبأخذ الجذر التربيعي نجد أن $\sqrt{S_{xx}} = 8.3666$. كما نجد أيضا أن $s = 0.27144$ ، من الجدول V نجد أن $t_{0.025} = 2.776$ وذلك من أجل 4 درجات حرية . وهكذا فإن 95% فترة الثقة المطلوبة لـ β هي :

$$\left(1.0257 - \frac{(2.776)(0.073685)}{8.3666} , 1.0257 + \frac{(2.776)(0.073685)}{8.3666} \right)$$

أو :

$$(1.00126 , 1.050148)$$

أي أن :

$$1.00125 < \beta < 1.050148$$

لاختيار الفرضية $H_0 : \beta = B_0$ ضد فرضية بديلة مناسبة ، فإننا نستخدم أيضا التوزيع t بـ (n − 2) درجة من الحرية ، وذلك لتجديد المنطقة الحرجة وبعد ذلك نبنى قرارنا على القيمة :

$$t = \frac{b - B_0}{s / \sqrt{S_{xx}}}$$

مثال (٨,٦)

باستخدام التقدير $b = 1.0257$ في المثال السابق اختبر الفرضية $\beta = 98$ ضد الفرضية $\beta > .98$ ، استخدم مستوى للمعنوية a 5 0.01

الحل

نلاحظ أن :

(١) فرض العدم هو : $H_0 : \beta = 0.98$

(٢) الفرضية البديلة هى : $H_1 : \beta > 0.98$

(٣) أن مستوى المعنوية هو : $\alpha = 0.01$

(٤) المنطقة الحرجة هى : $T > 3.747$

(٥) الإحصاء المحسوب من العينة هو : $T = \dfrac{1.0257 - 0.98}{0.073685 / 8.3666} = 5.189$

(٦) الاستنتاج : نرفض فرض العدم $H_0 : \beta = 0.98$ ونقبل الفرضية البديلة $H_1 : \beta > 0.98$ وذلك لأن قيمة الإحصاء $t = 5.189$ وقعت داخل المنطقة الحرجة $T > 3.747$.

يمكن بنفس الطريقة أن نبحث عن فترة الثقة ، وأن ندرس اختبار الثقة للمعامل λ إذا أخذنا بعين الاعتبار أن المتغير A طبيعى . من السهل التأكد من أن للإحصاء :

$$T = \frac{A - \lambda}{S\sqrt{\sum_{i=1}^{n} x_i^2 / n\, S_{xx}}}$$

توزيعاً من نوع t بـ $(n - 2)$ درجة من الحرية ، وبهذا الشكل نستنتج أن $(1 - \alpha)\, 100\%$ فترة الثقة للمعامل λ فى خط الانحدار $\mu_{Y|x} = \lambda + \beta x$ هى :

$$a = \frac{t_{\alpha/2} \cdot S\sqrt{\sum_{i=1}^{n} x_i^2}}{\sqrt{n.\ S_{x.x}}} < \lambda < a + \frac{t_{\alpha/2} \cdot S\sqrt{\sum_{i=1}^{n} x_i^2}}{\sqrt{n.\ S_{xx}}}$$

حيث يمثل $t_{\alpha/2}$ قيمة t بـ $(n - 2)$ درجة من الحرية .

لاختبار فرضية $H_0 : \lambda = \lambda_0$ ضد فرضية بديلة مناسبة ، يمكننا استخدام التوزيع t بـ $(n - 2)$ درجة حرية ، وذلك لتحديد المنطقة الحرجة ، وبعد ذلك نبنى قرارنا حول القيمة :

$$t = \frac{\lambda - \lambda_0}{S\sqrt{\sum_{i=1}^{n} x_i^2 / n\, S_{xx}}}$$

لنفرض أننا نهتم بتوقع Y من أجل القيمة $x = x_0$ عندئذ يمثل المقدار :

$$Y_0 = A + B x_0$$

تقديرا للوسط $\mu_{Y|x_0} = \lambda + \beta x_0$

وحيث إن : $E(\hat{Y}_0) = E(A + Bx_0) = \lambda + \beta x_0$

فهذا يعنى أن التقدير $\hat{Y}_0$ هو تقدير غير متحيز للوسط $\mu_{Y|x_0}$

بالتباين :

$$\sigma^2_{\hat{Y}_0} = \sigma^2_{A + \beta x_0} = \sigma^2_{\hat{Y} + B(x_0 - \bar{x})}$$

$$= \sigma^2 \left[\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right]$$

وهكذا فإن بإمكاننا إنشاء $(1 - \alpha)\ 100\%$ فترة ثقة للوسط $\mu_{Y|x_0}$ بواسطة الإحصاء :

$$T = \frac{\hat{Y}_0 - \mu_{Y|x_0}}{S \sqrt{\frac{1}{n} + [(x_0 - \bar{x})^2 / S_{xx}]}}$$

والذى يتوزع وفق توزيع t بـ $(n - 2)$ درجة من الحرية .

إن $(1 - \alpha)\ 100\%$ فترة الثقة للوسط $\mu_{Y|x_0}$ هى :

$$\hat{y}_0 - t_{\alpha/2}\, S \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}} < \mu_{Y|x_0} < \hat{y}_0 + t_{\alpha/2}\, S \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}}$$

حيث تمثل $t_{\alpha/2}$ قيمة التوزيع t بـ $(n - 2)$ درجة من الحرية .

مثال (٨,٧)

أوجد فى المثال (٨,٥) 95% فترة للوسط $\mu_{Y|x}$ أى وسط الصامد وذلك من أجل $x = 25\ kN/m^2$

الحل

نلاحظ من معادلة الانحدار أنه من أجل $x = 25$ فإن تقدير الوسط $\hat{y}_0$ هو :

$$\hat{y}_0 = 4.0888 + (1.0257)(25) = 29.73 \text{ kN/m}^2$$

إن 95% فترة ثقة لهذا التقدير هى :

$$\left(\hat{y}_0 - t_{0.025} S\sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}} , \hat{y}_0 + t_{0.025} S\sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}} \right)$$

أى :

$$\left((29.73 - (2.776)(0.27144)\sqrt{\frac{1}{6} + \frac{(25 - 16)^2}{70}} , \right.$$

$$\left. 29.739 + (2.776)(0.27144)\sqrt{\frac{1}{6} + \frac{(25 - 16)^2}{70}} \right)$$

$$= (28.863 , 30.606)$$

أى :

$$28.863 < \mu_{Y|x} < 30.606$$

لنفرض أننا نريد البدء بالصفات التوزيعية للفروقات بين الترتيبات المحسوبة من معادلة خط الانحدار فى العينات المتكررة والترتيب الحقيقى المقابل y_0 عند القيمة $x = x_0$ ، عندئذ يمكننا التفكير بالفرق $\hat{y}_0 - y_0$ كقيمة للمتغير العشوائى $\hat{Y}_0 - Y_0$ والذى يمكن البرهان بأن له توزيع طبيعى بالوسط :

$$\mu_{\hat{Y}_0 - Y_0} = E(\hat{Y}_0 - Y_0)$$

$$= E[(A + Bx_0) - (\lambda + \beta x_0 + E_0)]$$

$$= 0$$

والتباين :

$$\sigma^2_{\hat{Y}_0 - Y_0} = \sigma^2_{A + Bx_0 - E_0}$$

$$= \sigma^2_{\bar{y} + B(x_0 - \bar{x}) - E\eta}$$

$$= \sigma^2 \left[1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right]$$

وعندئذ يمكن الحصول على $(1 - \alpha)$ 100% فترة ثقة للقيمة الوحيدة المتنبئة y_0 من الإحصاء :

$$T = \frac{\hat{Y}_0 - Y_0}{S\sqrt{1 + \frac{1}{n} + (x_0 - \bar{x})^2/S_{xx}}}$$

والذى يتوزع وفقا للتوزيع t بـ $(n - 2)$ درجة من الحرية . وهكذا فإن $(1 - \alpha)$ 100% فترة الثقة للقيمة الوحيدة y_0 هى :

$$\hat{y}_0 - t_{\alpha/2} S\sqrt{1 + \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{S_{xx}}} < y_0 < \hat{y}_0 + t_{\alpha/2} S\sqrt{1 + \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{S_{xx}}}$$

حيث يمثل $t_{a/2}$ قيمة لـ t بـ $(n - 2)$ درجة من الحرية .

مثال (٨,٨)

أوجد فى المثال (٨,٧) 95% فترة للقيمة التقديرية الوحيدة لصامد المقص والموافقة لـ $x = 25\ KN/m^2$

الحل

لدينا $n = 6$, $x_0 = 25$, $\bar{x} = 16$, $\hat{y}_0 = 29.739$, $S_{xx} = 70$, $S = 0.27144$ ، وأخيراً

$t_{0.025} = 2.776$ من أجل 4 درجات حرية ، لذلك فإن الثقة لـ y_0 هى :

$$29.739 \pm (2.776)(0.27144)\sqrt{1 + \frac{1}{6} + \frac{(25 - 16)^2}{70}}$$

$$29.739 \pm 1.1487$$

أى :

$$(20.59 \, , \, 30.888) \ KN/m^2$$

(٨,٦) تحليل التباين Analysis of variance

تُعالجُ مسألة تحليل نوعية تقدير خط الانحدار غالبا من خلال إجراء يسمى بتحليل التباين ، حيث تجزأ وفق هذا الإجراء الاختلافات الكلية فى المتغير المرتبط y إلى عوامل تُراقب وتُعالج بشكل نظامى .

لنفرض أن لدينا n من المعلومات التجريبية على شكل أزواج (x_i , y_i) ، ولنفرض أيضا أنه قد تم تقدير خط الانحدار ، لقد أوضحنا فى الفقرة (٨,٣) عند تقديرنا لـ σ^2 أن :

$$S_{yy} = b \, S_{xy} + SSE$$

وهذا يعنى أننا جزأنا مجموع المربعات الكلى فى y إلى جزأين سنرمز لهذين الجزأين بالرمزين :

$$SST = S_{yy} \quad , \quad SSR = b \, S_{xy}$$

وعندئذ ستأخذ المتطابقة السابقة الشكل الجديد التالى :

$$SST = SSE + SSR$$

نسمى الجزء الأول فى الطرف الأيمن من العلاقة السابقة SSR بمجموع المربعات العائد لخط الانحدار ، وهذا المجموع يؤثر على مقدار الانحراف فى قيم y والموضحة بالنظام (فى هذه الحالة الخط المستقيم المسلم به) ، أما الجزء الثانى SSE فهو يمثل مجموع مربعات

الأخطاء ، والذى يؤثر على الاختلافات حول خط الانحدار ، وحيث إن $\frac{SSE}{\sigma^2}$, $\frac{SSR}{\sigma^2}$ لمتغيرين من نوع كاى – مربع بـ (1) , (n – 2) درجة حري على الترتيب لذا فإن $\frac{SST}{\sigma^2}$ تمثل قيمة لمتغير عشوائى من نوع كاى – مربع بـ (n – 1) درجة حرية . وعند اختبار فرضية $H_0 : \beta = 0$ ضد فرضية بديلة $H_1 : \beta \quad 0$ حيث تشير فرضية العدم إلى أن الاختلاف فى Y لا يُفسر بواسطة الخط المستقيم ، وإنما بواسطة التردد العشوائى أو بواسطة الخط فإننا نحسب :

$$f = \frac{SSR / 1}{SST / (n - 2)} = \frac{SSR}{s^2}$$

ثم نرفض H_0 عند مستوى المعنوية α إذا كان $f > f_\alpha(1 , n - 2)$. ونستنتج أن هناك مقدار اختلاف هام فى الإجابة المحسوبة بواسطة النظام المسلم به (أى دالة الخط المستقيم) . أما إذا وقع الإحصاء F فى منطقة القبول ، فإننا نستنتج أن المعلومات لا تؤثر بشكل كاف فى دعم النظام المسلم به .

وعند حل هذه المسألة نبدأ بحساب :

$$SST = S_{yy}$$

$$SSR = bS_{xy}.$$

وبعد ذلك نستخدم المتطابقة :

$$SSE = SST - SSR$$

ثم نقوم بتلخيص هذا فى جدول كالتالى ، (وهو ما يسمى عادة بجدول تحليل التباين) :

مصدر الاختلاف	مجموع المربعات	درجة الحرية	وسط التربيع	f المحسوبة
الانحدار	SSR	1	SSR	$\frac{SSR}{S^2}$
الخطأ	SSE	n – 2	$S^2 = \frac{SSE}{n - 2}$	
المجموع	SST	n – 1		

جدول (٨,١) أو جدول تحليل التباين من أجل $\beta = 0$

ملاحظة

لقد استخدمنا فى الفقرة (٨,٤) عند اختبار الفرضية $H_0 : \beta = \beta_0$ ضد الفرضية البديلة $H_1 : \beta \quad \beta_0$ الإحصاء :

$$T = \frac{B - \beta_0}{S / \sqrt{S_{xx}}}$$

حيث يتوزع T وفق توزيع t بـ $(n - 2)$ درجة حرية ، ولقد فرضنا فرض العدم H_0 عندما وقع $|T| > t_{\alpha/2}$ عند مستوى المعنوية α ، ومن المهم الإشارة إلى أنه فى الحالة التى تكون فيها $H_0 : \beta = 0$ ، فإن قيمة الإحصاء T تصبح :

$$t = \frac{b}{S / \sqrt{S_{xx}}}$$

وهى تتطابق مع تلك المذكورة فى الجدول (٨,١) ، أعنى أن فرض العدم يقرر الاختلاف فى y ناشىء عن الحظ . ويستخدم تحليل التباين التوزيع F أكثر من التوزيع t . غير أن الإجرائين متماثلان ويبدو هذا بوضوح بكتابة :

$$t^2 = \frac{b^2 S_{xx}}{S^2} = \frac{b \, S_{xy}}{S^2} = \frac{SSR}{S^2}$$

والمطابقة لقيمة f المستخدمة فى تحليل التباين . إن العلاقة الأساسية بين التوزيع t بـ v درجة من الحرية ، والتوزيع F بـ 1 , v درجات حرية يعطى بالعلاقة :

$$t^2_{\alpha/2} = f_{\alpha}(1 \, , \, v)$$

(٨,٧) القياسات المتكررة لـ y

The frequent measurement for y

من أجل الحصول على معلومات كمية تتعلق بملائمة النظام المستخدم ، فإن على المجرب أن يجرى ملاحظات متكررة لكل قيمة لـ x ، بينما لا يهم حدوث مثل هذه التكرارات عند تقدير كل من λ و β ، وبالحقيقة فإن وجود الملاحظات المتكررة تحت

تصرف المجرب تمكنه من إجراء اختبار أهمية يهدف إلى تحديد ما إذا كان النظام المستخدم ملائما أم لا .

لنختار عينة عشوائية مؤلفة من n ملاحظة ، وذلك باستخدام k قيمة مختلفة لـ x مثلاً $x_1, x_2, ..., x_k$ وبحيث تحوى العينة n_1 ملاحظة للمتغير Y_1 الموافق لـ x_1 ، n_2 ملاحظة للمتغير Y_2 الموافق لـ x_2 ، ... ، n_k ملاحظة للمتغير Y_k الموافق لـ x_k ، وبحيث يكون $n = \sum_{i=1}^{k} n_i$. لنفرض أن y_{ij} تمثل القيمة ذات الرقم j للمتغير العشوائى Y_i ، ولنفرض أن :

$$y_i = T_i = \sum_{j=1}^{n_i} y_{ij}$$

$$\overline{y}_i = \frac{T_i}{n_i}$$

لنفرض مثلا أن $n_4 = 3$ عندئذ تكون قياسات Y_4 الموافقة لـ $x = x_4$ هى y_{41}, y_{42}, y_{43} وسيكون أيضا :

$$T_4 = y_{41} + y_{42} + y_{43}$$

يتكون مجموع مربعات الأخطاء من جزئين يمثل الأول منهما الكمية الناشئة عن الاختلاف بين قيم Y والقيم المعطاة لـ x ، ويسمى هذا الجزء بنقصان تطابق الملائمة (lack of fit contribution) ، وهو يعكس اختلافاً عشوائياً مجرداً أو خطأ تجريبياً صرفاً ، بينما يمثل الجزء الثانى قياس للاختلاف النظامى الذى يحدث تحت شروط عالية الدرجة ، وفى حالتنا هذه فإن هذه الشروط فى x غير التداخل الخطى أو التداخل من الدرجة الأولى . نلاحظ أنه باختيارنا لنظام خطى ما ، فإننا نفرض أساسا أن الجزء الثانى غير موجود ولذلك فإن مجموع مربعات الأخطاء تنسب إلى خطأ عشوائى . وفى هذه الحالة يكون المقدار $S^2 = \frac{SSE}{n-2}$ ممثلا لتقدير ما غير متحيز لـ σ^2 . ومع ذلك إذا كان النظام غير منسجم بشكل كاف مع المعلومات فعندئذ يكون مجموع مربعات الأخطاء متضخما وتحدث تقديراً غير متحيز لـ σ^2 ، وسواء أكان النظام ملائما للمعلومات أو غير ملائم ، فإن بإمكان المجرب دائما الحصول على تقدير غير متحيز لـ σ^2 ، وذلك عندما تتكرر الملاحظات .

كما نلاحظ أن :

$$S_i^2 = \sum_{j=1}^{n} (y_{ij} - \bar{y})^2 , i = 1, 2, ..., k$$

وأنه من أجل كل القيم المختلفة الـ k لـ x فإن :

$$S^2 = \frac{\sum_{i=1}^{k} (n_i - 1) S_i^2}{n - k} = \frac{\sum_{i=1}^{k} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2}{n - k}$$

يمثل بسط S^2 قياسا للخطأ التجريبى الصرف ، هذا ويمكن فصل مجموع مربعات الأخطاء بطريقة حسابية إلى جزئين يمثل أحدهما الخطأ التجريبى الصرف ، كما يمثل الثانى النقصان فى الملائمة ، (lack of fit) وذلك على النحو التالى :

١ ـــ نبدأ بحساب مجموع مربعات الخطأ الصرف :

$$\sum_{i=1}^{k} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 = \sum_{i=1}^{k} \sum_{j=1}^{n_i} y_{ij}^2 - \sum_{i=1}^{k} \frac{T_i^2}{n_i}$$

ولمجموع المربعات السابق $n - k$ درجة حرية مرتبطة به . أما وسطه فهو يمثل تقديرنا غير المتحيز $\sigma^2 > S^2$.

٢ ـــ يطرح مجموع مربعات الخطأ الصرف من مجموع مربعات الأخطاء SSE ، فإننا نحصل على مجموع المربعات المعزو إلى النقصان فى الملائمة . هذا ويمكن الحصول على درجات حرية نقصان الملائمة بطرح بسيط $(n-2) - (n-k) = k-2$.

يمكن إجمال الحسابات المطلوبة لاختبار فرضية بواسطة القياسات المتكررة لـ y فى مسألة الانحدار بجدول كالتالى :

إن مفهوم النقصان فى الملائمة مهم جداً فى تطبيقات تحليل الانحدار . وفى الحقيقة فإن الحاجة إلى تصميم تجربة تقدم بيانا عن النقصان فى الملائمة تصبح أكثر من مسألة ، وتصبح التقنية الأساسية اللازمة لذلك أكثر تعقيداً وبالتأكيد لا يمكن للمجرب أن يكون متأكدا دائما من أن البناء المسلم به (فى هذه الحالة نظام الانحدار الخطى) صحيح أو حتى يمثل

كافيا . يوضح المثال التالى كيف أن مجموع مربعات الأخطاء ينقسم إلى جزأين يمثلان الخطأ الصرف والنقصان فى الملاءمة .

جدول (٨,٢) تحليل التباين من أجل القياسات المتكررة لـ y

مصدر الاختلاف	مجموع المربعات	درجات الحرية	وسط المربعات	الحساب
الانحدار	SSR	1	SSR	$\frac{SSR}{2}$
الخطأ	SSE	n − 2		
النقصان فى الملاءمة	SEE − (SSE الصرف)	k − 2	$\frac{SSE - (SSE \text{ الصرف})}{k-2}$	$\frac{SSE - (SSE \text{ الصرف})}{S^2(k-2)}$
الخطأ الصرف	SSE الصرف	n − k	$S^2 = \frac{SSE \text{ الصرف}}{n-k}$	
المجموع	SST	n − 1		

مثال (٨,٩)

بفرض أن كميات المعادن المستخرجة من مادة خاصة لدى تعرضها لفترات تنشيف ذات أطوال مختلفة هى كالتالى :

الجدول (٨,٣)

y بالغرامات		x بالساعات
14.2	13.1	4.4
11.5	09.0	4.5
11.5	10.4	4.8
14.8	13.8	5.5
15.1	12.7	5.7
12.7	09.9	5.9
16.5	13.8	6.3
15.7	16.4	6.9
16.9	17.6	7.5
17.2	18.3	7.8

١ ـــ فتش عن تقدير للنظام الخطى $\mu_{Y|x} = \lambda + \beta x$

٢ ـــ اختبر النقصان فى الملائمة .

الحل

لدينا $n_1 = n_2 = \ldots = n_{10} = 2$ لذلك فإن :

$$S_{yy} = \sum_{i=1}^{10} \sum_{j=1}^{2} y_{ij}^2 - \frac{(\sum_{i=1}^{10} \sum_{j=1}^{2} y_{ij})^2}{20}$$

$$= 4089.23 - \frac{(281.1)^2}{20}$$

$$= 4089.23 - 3950.8605$$

$$= 138.3695$$

$$S_{xx} = \sum_{i=1}^{10} n_i x_i^2 - \frac{(\sum_{i=1}^{10} n_i x_i)^2}{20}$$

$$= 2\,(364.59) - \frac{(118.6)^2}{20}$$

$$= 729.18 - 703.298$$

$$= 25.882$$

$$S_{xy} = \sum_{i=1}^{10} \sum_{j=1}^{2} x_i y_{ij} - \frac{(\sum_{i=1}^{10} n_i x_i)(\sum_{i=1}^{10} \sum_{j=1}^{2} y_{ij})}{20}$$

$$= 1714.62 - \frac{(118.6)\,(281.1)}{20}$$

$$= 47.697$$

$\bar{y} = 14.055 \;,\; \bar{x} = 5.93$

إن معاملى الانحدار هما :

$$b = \frac{S_{xy}}{S_{xx}} = \frac{47.697}{25.882} = 1.8429$$

ومنه :

$$a = \bar{y} - b\bar{x}$$
$$= 14.055 - (1.8429)(5.93)$$
$$= 3.1266$$

لذلك فإن تقدير خط الانحدار هو :

$$\hat{y} = a + bx = 3.1266 + 1.8429\,x$$

نتبع الطرق الاعتيادية لاختبار نقصان الملائمة حيث نلاحظ أن :

١ ـــ فرض العدم : الانحدار فى x خطى : H_0
٢ ـــ الفرض البديل : الانحدار فى x غير خطى : H_1
٣ ـــ لنفرض أن مستوى المعنوية هو $\alpha = 0.05$
٤ ـــ المنطقة الحرجة هى $F > 3.07$ بـ 10,8 درجات حرية .

٥ ـــ الحسابات : لدينا

$$SST = S_{yy} = 138.3695$$

$$SSR = b\,S_{xy} = (1.8429)(47.697) = 87.9$$

$$SSE = S_{yy} - b\,S_{xy} = 138.3695 - 87.9 = 50.47$$

ولحساب مجموع مربعات الخطأ الصرف فإننا نكتب أولا :

x_1	= 4.4	T_1	= 27.3
x_2	= 4.5	T_2	= 20.5
x_3	= 4.8	T_3	= 21.9
x_4	= 5.5	T_4	= 28.6
x_5	= 5.7	T_5	= 27.8
x_6	= 5.9	T_6	= 27.6
x_7	= 6.3	T_7	= 30.3
x_8	= 6.9	T_8	= 32.1
x_9	= 7.5	T_9	= 34.5
x_{10}	= 7.8	T_{10}	= 35.5

لذلك فإن :

$$(\text{SSE الصرف}) = 4089.23 - \frac{(27.3)^2 + (20.5)^2 + (21.9)^2 + \ldots + (35.5)^2}{2}$$

$$= 4089.23 - 4072.86 = 16.375$$

يمكن إجمال النتائج السابقة بجدول كالتالى :

جدول (٨,٦) تحليل التباين فى فترات التنشيف المختلفة

مصدر الاختلاف	مجموع المربعات	درجات الحرية	وسط المجموع	حساب f
الانحدار	87.900	01	87.9000	53.68
الخطأ	50.470	18		
نقصان الملائمة	34.095	08	04.2619	02.60
الخطأ الصرف	16.375	10	01.6375	
المجموع	**138.37**	**19**	**93.7994**	**56.28**

٦ ــ الاستنتاج : إن تجزئة مجموع الاختلافات الكلية بهذه الطريقة يُظهر اختلافاً مهماً مرتبطاً بالنظام الخطى ، كما تُظهر كمية من الاختلافات غير المهمة والمنسوبة إلى النقصان فى الملائمة ، وهكذا فإن المعلومات التجريبية لا تقترح الحاجة إلى افتراض شروط أعلى من الدرجة الأولى فى النظام ، ونقبل بذلك فرض العدم .

(٨,٨) الارتباط The correlation

لقد فرضنا حتى الآن أننا نقوم بملاحظة المتغير المستقل x ، ولذلك فهذا المتغير ليس عشوائيا ، وفى الحقيقة فإنه يسمى فى هذا السياق بمتغير رياضى . وهو قياس فى العملية العينية بخطا مهمل . ومن الأجدر فى كثير من تطبيقات الانحدار التطبيقية الفرض بأن كلا من X , Y يمثل متغيراً عشوائياً ، وأن القياسات $\{(x_i , y_i) : i = 1,2, ..., n\}$ تمثل ملاحظات من دالة الكثافة المشتركة $f(x , y)$. ففى علم الآثار مثلا يمكن الفرض بأن قياسين لنوع معين من عظام جسم كائن بالغ هما متغيران عشوائيان يتبعان توزيعا ذى بعدين .

يفترض عادة أن التوزيع الشرطى $f(y|x)$ للمتغير Y من أجل قيم ثابتة لـ X هو توزيع طبيعى بالوسط $\mu_{Y|x} = \lambda + \beta x$ ، والتباين $\sigma^2_{Y|x} = \sigma^2$ ، وأن X يتوزع بطريقة مماثلة وفقا للتوزيع الطبيعى بالوسط μ_x ، والتباين σ^2_x ، وأن الكثافة المشتركة لـ X و Y تعطى بالعلاقة :

$$f(x,y) = n(y|x : \lambda + \beta x, \sigma) . n(x ; \mu_x , \sigma_x)$$

$$= \frac{1}{2\pi\ \sigma_x \sigma} \exp\left\{ -(\frac{1}{2}) \quad [(\frac{y - (\lambda + \beta x)}{\sigma})^2 + (\frac{x - \mu_x}{\sigma_x})^2] \right\}$$

حيث إن $-\infty < x < \infty$ و $-\infty < y < \infty$ وبكتابة المتغير Y على النحو :

$$Y = \lambda + \beta X + E$$

حيث يمثل X هنا متغيراً عشوائياً مستقلاً عن الخطأ العشوائى E ، فإننا نجد أن :

$$\mu_Y = \lambda + \beta \mu_x$$

$$\sigma^2_Y = \sigma^2 + \beta^2 \sigma^2_x$$

وبالتبديل فى عبارة f(x , y) السابقة ، فإننا نحصل على توزيع طبيعى ذى بعدين :

$$f(x, y) = \frac{1}{2\pi\,\sigma_x\sigma_Y\sqrt{1-\rho^2}} \exp \frac{-1}{2(1-\rho^2)} \left[\left(\frac{X-\mu_x}{\sigma_x}\right)^2 - 2\rho\left(\frac{X-\mu_x}{\sigma_x}\right)\left(\frac{Y-\mu_Y}{\sigma_Y}\right) + \left(\frac{Y-\mu_Y}{\sigma_Y}\right)^2 \right]$$

حيث إن $-\infty < x < \infty \,,\, -\infty < y < \infty \,,\, \rho = 1 - \frac{\sigma^2}{\sigma_Y^2} = \beta\,\frac{\sigma_x^2}{\sigma_Y^2}$

يسمى الثابت ρ بمعامل الارتباط . وهو يلعب دوراً هاماً فى عدد من مسائل تحليل البيانات ذات البعدين . ومن المهم للقارىء أن يفهم التفسير الفيزيائى لمعامل الارتباط والاختلاف بين الارتباط والانحدار . ويظل لتعبير الانحدار هنا معنى ، وبالحقيقة فإن الخط المستقيم $\mu_{Y|x} = \lambda + \beta X$ يظل اسمه خط انحدار كما فى السابق وأن تقديرات λ , β تطابق التقديرات الواردة فى الفقرة (٨,٢) ، ويأخذ ρ القيمة صفر عندما يكون $\beta = 0$ ، ويحدث هذا عندما لا يكون هناك انحداراً خطياً ، وهذا يعنى أن خطأ الانحدار أفقى ، وأن معرفتنا بـ X لا تفيد فى التنبؤ بقيم Y . وبما أن $\sigma_Y^2 = \sigma^2$ لذا فإننا نجد أن $-1 \le \rho \le 1$ ، ويأخذ ρ القيمتين ± 1 عندما يكون $\sigma^2 = 0$. وهذا يعنى وجود علاقة خطية كاملة بين المتغيّرين ، وهكذا فإن القيمة $\rho = 1$ تدل ضمنا على وجود علاقة خطية كاملة بميل موجب ، بينما تشير القيمة $\rho = -1$ على وجود علاقة خطية كاملة بميل سالب .

هذا ويمكن القول بأن التقديرات العينية لـ ρ القريبة من الواحد تقتضى إما ارتباط جيداً أو علاقة خطية بين X و Y ، بينما تشير القيم القريبة من الصفر إما إلى ارتباط بسيط أو عدمه .

يجب التأكيد على أن قيمة ρ المحسوبة تقيس فى (أية حالة) درجة العلاقة بالنسبة لنوع المعادلة المفروضة . فإذا فرضنا معادلة خطية ونتج أن قيمة ρ تقترب من الصفر ، فهذا يعنى أنه لا يوجد تقريبا علاقة خطية بين المتغيرين ، ولكن هذا لا يعنى أنه لايوجد

علاقة بين المتغيرين على الإطلاق ، حيث إنه قد يكون هناك بالفعل علاقة كبيرة غير خطية بين هذين المتغيرين . وبصورة أخرى فإن معامل الارتباط يقيس مدى جودة توفيق المعادلة المفترضة للبيانات . إن مصطلح معامل الارتباط يستخدم ليعنى الارتباط الخطى مالم نُشير إلى خلاف ذلك . ويجب إيضاح أن وجود معامل ارتباط مرتفع ، أى يقترب من 1 أو 1 − لا يعنى وجود علاقة تبعية مباشرة بين المتغيرين ، فقد يكون هناك معامل ارتباط مرتفع بين عدد الذين تزوجوا وعدد الأشجار المثمرة ، مثل هذه الأمثلة يشار إليها بأنها ارتباط لا معنى له أو ارتباط زائف .

للحصول على تقدير عينى لمعامل الارتباط ρ فإننا نعود قليلا إلى الفقرة (٨,٣) لنجد أن مجموع مربعات الأخطاء تعطى بالعلاقة :

$$SSE = S_{yy} - b\,S_{xy}$$

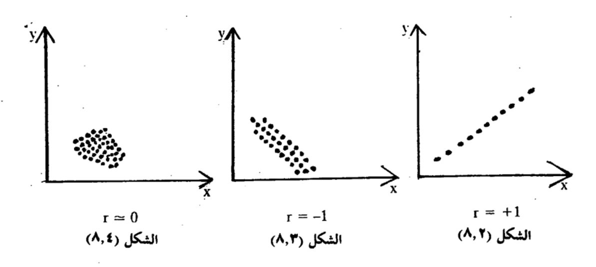
وبقسمة طرفى المعادلة الأخيرة على S_{yy} وبتبديل $b = \frac{S_{xy}}{S_{xx}}$ فإننا نحصل على العلاقة :

$$b^2\,\frac{S_{xx}}{S_{yy}} = 1 - \frac{SSE}{S_{yy}}$$

إن قيمة $b^2\,\frac{S_{xx}}{S_{yy}}$ تساوى الصفر عندما يكون $b = 0$ و يحدث هذا عندما لا يكون هناك علاقة خطية بين نقاط العينة . وبما أن $S_{yy} \geq SSE$ فإننا نستنتج أن $b^2\,\frac{S_{xx}}{S_{yy}}$ تقع بين الصفر والواحد ، لذلك فإن $b\sqrt{S_{xx}/S_{yy}}$ تتردد بين -1 , $+1$ وتوافق القيم السالبة مستقيمات بميول سالبة ، كما توافق القيم الموجبة مستقيمات بميول موجبة ، وسيأخذ هذا المقدار القيمة 1 أو -1 عندما يكون $SSE = 0$. وتوضع فى هذه الحالة جميع قيم العينة على خط مستقيم . لذلك فإن العلاقة الخطية الكاملة تحدث بين X و Y عندما يكون :

$$b\sqrt{S_{xx}/S_{yy}} = \pm 1$$

تمثل الأشكال (٨,٢) ، (٨,٣) ، (٨,٤) على الترتيب أمثلة على $r = +1$ و r تساوى تقريبا -1 و r تساوى تقريبا صفراً .



$r \simeq 0$ | $r = -1$ | $r = +1$

الشكل (٨,٤) الشكل (٨,٣) الشكل (٨,٢)

من الواضح أنه يمكن استخدام الكمية $b\sqrt{S_{xx}/S_{yy}}$ والتي سنرمز لها من الآن فصاعدا بالرمز r كتقدير لمعامل الارتباط P .

تعريف معامل الارتباط The correlation coefficienf

إن القياس ρ لدرجة العلاقة الخطية بين متغيرين X و Y تقدر بواسطة معامل الارتباط العينى r حيث :

$$r = b\sqrt{\frac{S_{xx}}{S_{yy}}} = \frac{S_{xy}}{\sqrt{S_{xx}\ S_{yy}}}$$

وتنحصر قيمة r بين 1 - , 1 + ، ويجب أن نكون حريصين فى تفسيراتنا ، فمثلا قيم r المساوية لـ 0.2 و 0.4 تعنى فقط أن لدينا إرتباطين موجبين وأحدهما أقوى من الآخر إلى حد ما ، ومن الخطأ أن نستنتج أن $r = 0.4$ تشير إلى وجود علاقة خطية أفضل بمرتين من التى تشير إليها القيمة $r = 0.2$ ، ومن ناحية أخرى إذا افترضنا r^2 فإن $r^2\%$ (100) من الاختلافات فى قيم Y يمكن أن تحسب بواسطة العلاقة الخطية بالمتغير X . وهكذا فإن الارتباط 0.4 يعنى أن 16% من الاختلافات فى المتغير Y تحسب بواسطة الفروق فى المتغير X .

مثال (٨,١٠)

إذا علمت أن أطوال ثمانية آباء وأطوال أكبر أولادهم معطاة بالجدول (٨,٥) :

جدول (٨,٥)

x طول الأب بالانش	y طول أكبر أولاده
63	65
64	67
70	69
72	70
65	64
67	68
68	71
66	63

أ ــ أوجد من أجل هذه المعلومات خط انحدار y على x .
ب ــ أوجد تقديراً لطول أكبر الأولاد وذلك من أجل طول الأب 69 انش .
جـ ــ ارسم خط الانحدار .
د ــ احسب معامل الارتباط العينى r ثم فسر هذا الارتباط .

الحل

أولا : من أجل إيجاد خط الانحدار نلاحظ أن :

الجدول (٨.٨)

x	y	xy	x^2	y^2	$\hat{y}$	$y - \hat{y}$	$(y - \hat{y})^2$
63	65	4095	3969	4225	64.55	0.45	0.2025
64	67	4288	4096	4489	65.21	1.79	3.2041
70	69	4830	4900	4761	69.20	−0.20	0.0400
72	70	5040	5184	4900	70.53	−0.53	0.2809
65	64	4160	4225	4096	65.88	−1.88	3.5344
67	68	4556	4489	4624	67.21	.78	0.6241
68	71	4828	4624	5041	76.87	3.13	9.7969
66	63	4158	4356	3969	66.54	−3.54	12.5316
535	537	35955	35843	36105			30.2145

كما نلاحظ أن $\bar{x} = \frac{535}{8} = 66.875$, $\bar{y} = \frac{537}{8} = 67.125$

ومن المعلوم أن معادلة انحدار Y على X هى :

$$\hat{y} = a + bx$$

كما نلاحظ أن :

$$b = \frac{\sum_{i=1}^{8} x_i y_i - n\bar{x}\,\bar{y}}{\sum_{i=1}^{8} x_i^2 - n.\,\bar{x}^2}$$

$$= \frac{35955 - 8(66.875)\,(67.125)}{35843 - 8(66.875)^2}$$

وأن :

$$a = \bar{y} - b\bar{x} = 67.125 - (0.6647)\,(66.875)$$
$$= 22.6732$$

لذلك فإن معادلة الانحدار المطلوبة هى :

$$\hat{y} = 22.6732 - 6647\,x$$

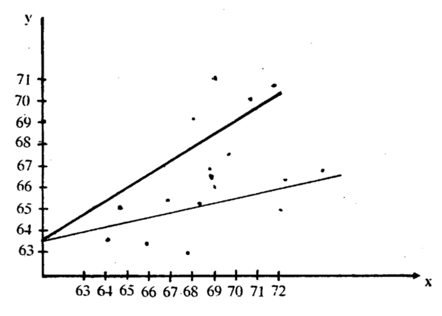
ثانيا : باستخدام معادلة الانحدار من أجل $x = 69$ ، فإننا نجد أن :

$$\hat{y} = 22.6732 + 6647(69) = 68.5373$$

ثالثا : من أجل رسم الانحدار السابق ، فإننا نحتاج إلى نقطتين فقط ، نلاحظ مثلاً أنه من أجل :

$x = 62$	أن	$\hat{y} = 63.8846$
$x = 69$	أن	$\hat{y} = 68.5375$

وسنحصل عندئذ على خط الانحدار $\hat{y} = 22.6732 + 0.6647x$ بالوصل بين النقطتين $M(62,\ 63.8846)$, $N(69,\ 68.5375)$. يوضح الشكل (٨,٥) مخطط الانتشار وخط الانحدار المطلوب .



الشكل (٨,٥)

نلاحظ أيضاً أن :

$$S_{xx} = \sum_{i=1}^{8} x_i^2 - \frac{(\sum_{i=1}^{8} x_i)^2}{8} = 64.875$$

$$S_{xx} = \sum_{i=1}^{8} y_i^2 - \frac{(\sum_{i=1}^{8} y_i)^2}{8} = 58.875$$

$$S_{xy} = \sum_{i=1}^{n} x_i y_i - \frac{(\sum_{i=1}^{8} x_i)(\sum_{i=1}^{8} y_i)}{8} = 43.125$$

لذلك فإن :

$$r = \frac{S_{xy}}{\sqrt{S_{xx}\ S_{yy}}} = \frac{34.125}{\sqrt{(64.875)\ (58.875)}}$$

ومنه :

$$r = 6977$$

يشـــير معامل الارتباط الســـابق إلى وجود علاقة خطية بين X و Y ، وحيث إن $r^2 = 0.4869$ ، لذلك فإن بإمكاننا القول بأن 48.69% تقريبا من الاختلافات فى قيم y تحسب بواسطة علاقة انحدار Y على X .

لاختبار الفرضية $H_0 : \rho = \rho_0$ ضد الفرضية البديلة $H_1 : \rho \quad \rho_0$ ، فإننا نلاحظ أنه من ملاحظات التوزيع الطبيعى ذى البعدين فإن الكمية $\frac{1}{2} \ln \frac{1+r}{1-r}$ تمثل قيمة متغير عشوائى يتبع تقريبا التوزيع الطبيعى بالوسط $\frac{1}{2} \ln \frac{1+\rho}{1-\rho}$ والتباين $\frac{1}{n-3}$ ، وهكذا فإن إجراء الاختبار هو فى حساب :

$$z = \frac{\sqrt{n-3}}{2} \left[\ln \left(\frac{1+r}{1-r} \right) - \ln \left(\frac{1+\rho_0}{1-\rho_0} \right) \right]$$

$$= \frac{\sqrt{n-3}}{n} \ln \left[\frac{(1+r)(1-\rho_0)}{(1-r)(1+\rho_0)} \right]$$

ومقارنته مع النقاط الحرجة للتوزيع الطبيعى المعيارى .

مثال (٨,١١)

اختبر فى المثال السابق الفرضية فى أنه ليس هناك ارتباط خطى بين المتغيرين . استخدم لذلك مستوى معنوية $\alpha = 0.05$.

الحل

١ ــ إن فرض العدم هو $H_0 : \rho = 0$

٢ ــ إن الفرض البديل هو $H_1 : \rho \neq 0$

٣ ــ $\alpha = 0.05$

٤ ــ إن المنطقة الحرجة هى $Z < 1.96, Z < -1.96$

٥ ــ الحسابات

$$z = \frac{\sqrt{5}}{2} \ln \left[\frac{1.6978}{0.3022} \right] = 1.9297$$

٦ ــ الاستنتاج : إننا نقبل الفرضية بعدم وجود علاقة خطية .



Barr, D.R. and **Zehna, P.W.** (1971) *Probability*, California: Brooks/Cole Publishing Co.

Bowker, A.H. and **Lieberman, G.J.** (1972) *Engineering Statistics*, 3rd ed., Englewood Cliff, Prentice Hall Inc.

Feller, William (1966) *An Introduction to Probability and its Applications*, New York: John Wiley, **Vol. 1,** 1968, 3rd ed., and **Vol. 2.**

Fisz, Marek (1963) *Probability Theory and Mathematical Statistics*, 3rd ed., New York: John Wiley.

Freund, J.E. (1971) *Mathematical Statistics*, 2nd ed., Englewood Cliffs, Prentice-Hall Inc.

Guttman, I.S.S. Wilks and **Hunter, J.S.** (1971) *Introductory Engineering Statistics*, 2nd ed., New York: John Wiley and Sons.

Hawkins, C.A. and **Weber, J.E.** (1980) *Statistical Analysis*, New York: Harper and Row.

Larson, Harold J. (1973) *Introduction to the Theory of Statistics*, New York: John Wiley.

McClave, J.T. and **Dietrich, F.H.** (1979) *Statistics*, San Francisco: Dellen Publishing Co.

Mendenhall, W., Scheaffer, R.L. and **Wackerly, D.D.** (1981) *Mathematical Statistics with Applications*, 2nd ed., Boston: Duxbury Press.

Miller, I. and **Freund, J.E.** (1977). *Probability and Statistics for Engineers*, 2nd ed., Englewood Cliffs, Prentice-Hall, Inc.

Rao, C.R. (1973) *Linear Statistical Inference and its Applications*, 2nd ed., New York: John Wiley.

Scheaffer, Richard, L. and *McClave, James T.* (1982) *Statistics for Engineers*, Boston: Duxbury Press.

Stephens, M.A. (1974) EDF statistics for goodness and fit and some comparisons, *Journal of Am. Sta Assn.*, **Vol. 69,** no. 347, pp. 730-737.

Walpole, R.E. and **Myers, R.H.** (1978) *Probability and Statistics for Engineers and Scientists*, 2nd ed., New York: Macmillan Publishing Co.

Zehna, Peter W. (1970) *Probability Distributions and Statistics*. Boston: Allyn and Bacon, Inc.

الملاحق

■ جدول I : المربعات والجذور التربيعية ■ جدول II : مجموع الاحتمال الحداني $\sum_{x=0}^{r} b(x;n.p)$ ■ جدول III : مجموع الاحتمال البواسوني $\sum_{x=0}^{r} p(x;\mu)$ ■ جدول IV : المساحة تحت المنحنى الطبيعي ■ جدول V : القيم الحرجة في توزيع t ■ جدول VI : القيم الحرجة في توزيع كاي مربع ■ جدول VII : القيم الحرجة في توزيع f ■ جدول VIII : عوامل التحمل في التوزيع الطبيعي .



جدول I : المربعات والجذور التربيعية

n	n^2	$\sqrt{n}$	$\sqrt{10n}$	n	n^2	$\sqrt{n}$	$\sqrt{10n}$
1.0	1.00	1.000	3.162	5.5	30.25	2.345	7.416
1.1	1.21	1.049	3.317	5.6	31.36	2.366	7.483
1.2	1.44	1.095	3.464	5.7	32.49	2.387	7.550
1.3	1.69	1.140	3.606	5.8	33.64	2.408	7.616
1.4	1.96	1.183	3.742	5.9	34.81	2.429	7.681
1.5	2.25	1.225	3.873	6.0	36.00	2.449	7.746
1.6	2.56	1.265	4.000	6.1	37.21	2.470	7.810
1.7	2.89	1.304	4.123	6.2	38.44	2.490	7.874
1.8	3.24	1.342	4.243	6.3	39.69	2.510	7.937
1.9	3.61	1.378	4.359	6.4	40.96	2.530	8.000
2.0	4.00	1.414	4.472	6.5	42.25	2.550	8.062
2.1	4.41	1.449	4.583	6.6	43.56	2.569	8.124
2.2	4.84	1.483	4.690	6.7	44.89	2.588	8.185
2.3	5.29	1.517	4.796	6.8	46.24	2.608	8.246
2.4	5.76	1.549	4.899	6.9	47.61	2.627	8.307
2.5	6.25	1.581	5.000	7.0	49.00	2.646	8.367
2.6	6.76	1.612	5.099	7.1	50.41	2.6[illegible]	8.426
2.7	7.29	1.643	5.196	7.2	51.84	2.68[illegible]	8.485
2.8	7.84	1.673	5.292	7.3	53.29	2.702	8.544
2.9	8.41	1.703	5.385	7.4	54.76	2.720	8.602
3.0	9.00	1.732	5.477	7.5	56.25	2.739	8.660
3.1	9.61	1.761	5.568	7.6	57.76	2.757	8.718
3.2	10.24	1.789	5.657	7.7	59.29	2.775	8.775
3.3	10.89	1.817	5.745	7.8	60.84	2.793	8.832
3.4	11.56	1.844	5.831	7.9	62.41	2.811	8.888
3.5	12.25	1.871	5.916	8.0	64.00	2.828	8.944
3.6	12.96	1.897	6.000	8.1	65.61	2.846	9.000
3.7	13.69	1.924	6.083	8.2	67.24	2.864	9.055
3.8	14.44	1.949	6.164	8.3	68.89	2.881	9.110
3.9	15.21	1.975	6.245	8.4	70.56	2.898	9.165
4.0	16.00	2.000	6.325	8.5	72.25	2.915	9.220
4.1	16.81	2.025	6.403	8.6	73.96	2.933	9.274
4.2	17.64	2.049	6.481	8.7	75.69	2.950	9.327
4.3	18.49	2.074	6.557	8.8	77.44	2.966	9.381
4.4	19.36	2.098	6.633	8.9	79.21	2.983	9.434
4.5	20.25	2.121	6.708	9.0	81.00	3.000	9.487
4.6	21.16	2.145	6.782	9.1	82.81	3.017	9.539
4.7	22.09	2.168	6.856	9.2	84.64	3.033	9.592
4.8	23.04	2.191	6.928	9.3	86.49	3.050	9.644
4.9	24.01	2.214	7.000	9.4	88.36	3.066	9.695
5.0	25.00	2.236	7.071	9.5	90.25	3.082	9.747
5.1	26.01	2.258	7.141	9.6	92.16	3.098	9.798
5.2	27.04	2.280	7.211	9.7	94.09	3.114	9.849
5.3	28.09	2.302	7.280	9.8	96.04	3.130	9.899
5.4	29.16	2.324	7.348	9.9	98.01	3.146	9.950

جدول II : مجموع الاحتمال الحداني $\sum_{x=0}^{r} b(x;n,p)$

n	r	p = 0.10	0.20	0.25	0.30	0.40	0.50	0.60	0.70	0.80	0.90
5	0	0.5905	0.3277	0.2373	0.1681	0.0778	0.0312	0.0102	0.0024	0.0003	0.0000
	1	0.9185	0.7373	0.6328	0.5282	0.3370	0.1875	0.0870	0.0308	0.0067	0.0005
	2	0.9914	0.9421	0.8965	0.8369	0.6826	0.5000	0.3174	0.1631	0.0579	0.0086
	3	0.9995	0.9933	0.9844	0.9692	0.9130	0.8125	0.6630	0.4718	0.2627	0.0815
	4	1.0000	0.9997	0.9990	0.9976	0.9898	0.9688	0.9222	0.8319	0.6723	0.4095
	5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
10	0	0.3487	0.1074	0.0563	0.0282	0.0060	0.0010	0.0001	0.0000	0.0000	0.0000
	1	0.7361	0.3758	0.2440	0.1493	0.0464	0.0107	0.0017	0.0001	0.0000	0.0000
	2	0.9298	0.6778	0.5256	0.3828	0.1673	0.0547	0.0123	0.0016	0.0001	0.0000
	3	0.9872	0.8791	0.7759	0.6496	0.3823	0.1719	0.0548	0.0106	0.0009	0.0000
	4	0.9984	0.9672	0.9219	0.8497	0.6331	0.3770	0.1662	0.0474	0.0064	0.0002
	5	0.9999	0.9936	0.9803	0.9527	0.8338	0.6230	0.3669	0.1503	0.0328	0.0016
	6	1.0000	0.9991	0.9965	0.9894	0.9452	0.8281	0.6177	0.3504	0.1209	0.0128
	7	1.0000	0.9999	0.9996	0.9984	0.9877	0.9453	0.8327	0.6172	0.3222	0.0702
	8	1.0000	1.0000	1.0000	0.9999	0.9983	0.9893	0.9536	0.8507	0.6242	0.2639
	9	1.0000	1.0000	1.0000	1.0000	0.9999	0.9990	0.9940	0.9718	0.8926	0.6513
	10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
15	0	0.2059	0.0352	0.0134	0.0047	0.0005	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.5490	0.1671	0.0802	0.0353	0.0052	0.0005	0.0000	0.0000	0.0000	0.0000
	2	0.8159	0.3980	0.2361	0.1268	0.0271	0.0037	0.0003	0.0000	0.0000	0.0000
	3	0.9444	0.6482	0.4613	0.2969	0.0905	0.0176	0.0019	0.0001	0.0000	0.0000
	4	0.9873	0.8358	0.6865	0.5155	0.2173	0.0592	0.0094	0.0007	0.0000	0.0000
	5	0.9978	0.9389	0.8516	0.7216	0.4032	0.1509	0.0338	0.0037	0.0001	0.0000
	6	0.9997	0.9819	0.9434	0.8689	0.6098	0.3036	0.0951	0.0152	0.0008	[illegible]
	7	1.0000	0.9958	0.9827	0.9500	0.7869	0.5000	0.2131	0.0500	0.0042	[illegible]
	8	1.0000	0.9992	0.9958	0.9848	0.9050	0.6964	0.3902	0.1311	0.0181	0.0003
	9	1.0000	0.9999	0.9992	0.9963	0.9662	0.8491	0.5968	0.2784	0.0611	0.0023
	10	1.0000	1.0000	0.9999	0.9993	0.9907	0.9408	0.7827	0.4845	0.1642	0.0127
	11	1.0000	1.0000	1.0000	0.9999	0.9981	0.9824	0.9095	0.7031	0.3518	0.0556
	12	1.0000	1.0000	1.0000	1.0000	0.9997	0.9963	0.9729	0.8732	0.6020	0.1841
	13	1.0000	1.0000	1.0000	1.0000	1.0000	0.9995	0.9948	0.9647	0.8329	0.4510
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9995	0.9953	0.9648	0.7941
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
20	0	0.1216	0.0115	0.0032	0.0008	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.3917	0.0692	0.0243	0.0076	0.0005	0.0000	0.0000	0.0000	0.0000	0.0000
	2	0.6769	0.2061	0.0913	0.0355	0.0036	0.0002	0.0000	0.0000	0.0000	0.0000
	3	0.8670	0.4114	0.2252	0.1071	0.0160	0.0013	0.0001	0.0000	0.0000	0.0000
	4	0.9568	0.6296	0.4148	0.2375	0.0510	0.0059	0.0003	0.0000	0.0000	0.0000
	5	0.9887	0.8042	0.6172	0.4164	0.1256	0.0207	0.0016	0.0000	0.0000	0.0000
	6	0.9976	0.9133	0.7858	0.6080	0.2500	0.0577	0.0065	0.0003	0.0000	0.0000
	7	0.9996	0.9679	0.8982	0.7723	0.4159	0.1316	0.0210	0.0013	0.0000	0.0000
	8	0.9999	0.9900	0.9591	0.8867	0.5956	0.2517	0.0565	0.0051	0.0001	0.0000
	9	1.0000	0.9974	0.9861	0.9520	0.7553	0.4119	0.1275	0.0171	0.0006	0.0000
	10	1.0000	0.9994	0.9961	0.9829	0.8725	0.5881	0.2447	0.0480	0.0026	0.0000
	11	1.0000	0.9999	0.9991	0.9949	0.9435	0.7483	0.4044	0.1133	0.0100	0.0001
	12	1.0000	1.0000	0.9998	0.9987	0.9790	0.8684	0.5841	0.2277	0.0321	0.0004
	13	1.0000	1.0000	1.0000	0.9997	0.9935	0.9423	0.7500	0.3920	0.0867	0.0024
	14	1.0000	1.0000	1.0000	1.0000	0.9984	0.9793	0.8744	0.5836	0.1958	0.0113
	15	1.0000	1.0000	1.0000	1.0000	0.9997	0.9941	0.9490	0.7625	0.3704	0.0432
	16	1.0000	1.0000	1.0000	1.0000	1.0000	0.9987	0.9840	0.8929	0.5886	0.1330
	17	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9964	0.9645	0.7939	0.3231
	18	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9995	0.9924	0.9308	0.6083
	19	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9992	0.9885	0.8784
	20	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

جدول III : مجموع الاحتمال البواسوني $\sum_{x=0}^{r} P(x\,;\,\mu)$

r	μ 0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0	0.9048	0.8187	0.7408	0.6730	0.6065	0.5488	0.4966	0.4493	0.4066
1	0.9953	0.9825	0.9631	0.9384	0.9098	0.8781	0.8442	0.8088	0.7725
2	0.9998	0.9989	0.9964	0.9921	0.9856	0.9769	0.9659	0.9526	0.9371
3	1.0000	0.9999	0.9997	0.9992	0.9982	0.9966	0.9942	0.9909	0.9865
4		1.0000	1.0000	0.9999	0.9998	0.9996	0.9992	0.9986	0.9977
5				1.0000	1.0000	1.0000	0.999[illegible]	0.9998	0.9997
6							1.0000	1.0000	1.0000

جدول III : مجموع الاحتمال البواسوني (تابع)

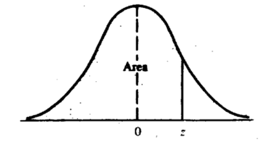
r	μ 1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
0	0.3679	0.2231	0.1353	0.0821	0.0498	0.0302	0.0183	0.0111	0.0067
1	0.7358	0.5578	0.4060	0.2873	0.1991	0.1359	0.0916	0.0611	0.0404
2	0.9197	0.8088	0.6767	0.5438	0.4232	0.3208	0.2381	0.1736	0.1247
3	0.9810	0.9344	0.8571	0.7576	0.6472	0.5366	0.4335	0.3423	0.2650
4	0.9963	0.9814	0.9473	0.8912	0.8153	0.7254	0.6288	0.5321	0.4405
5	0.9994	0.9955	0.9834	0.9580	0.9161	0.8576	0.7851	0.7029	0.6160
6	0.9999	0.9991	0.9955	0.9858	0.9665	0.9347	0.8893	0.8311	0.7622
7	1.0000	0.9998	0.9989	0.9958	0.9881	0.9733	0.9489	0.9134	0.8666
8		1.0000	0.9998	0.9989	0.9962	0.9901	0.9786	0.9597	0.9319
9			1.0000	0.9997	0.9989	0.9967	0.9919	0.9829	0.9682
10				0.9999	0.9997	0.9990	0.9972	0.9933	0.9863
11				1.0000	0.9999	0.9997	0.9991	0.9976	0.9945
12					1.0000	0.9999	0.9997	0.9992	0.9980
13						1.0000	0.9999	0.9997	0.9993
14							1.0000	0.9999	0.9998
15								1.0000	0.9999
16									1.0000

جدول III : مجموع الاحتمال البواسوني (تابع)

	μ								
r	5.5	6.0	6.5	7.0	7.5	8.0	8.5	9.0	9.5
0	0.0041	0.0025	0.0015	0.0009	0.0006	0.0003	0.0002	0.0001	0.0001
1	0.0266	0.0174	0.0113	0.0073	0.0047	0.0030	0.0019	0.0012	0.0008
2	0.0884	0.0620	0.0430	0.0296	0.0203	0.0138	0.0093	0.0062	0.0042
3	0.2017	0.1512	0.1118	0.0818	0.0591	0.0424	0.0301	0.0212	0.0149
4	0.3575	0.2851	0.2237	0.1730	0.1321	0.0996	0.0744	0.0550	0.0403
5	0.5289	0.4457	0.3690	0.3007	0.2414	0.1912	0.1496	0.115	0.0885
6	0.6860	0.6063	0.5265	0.4497	0.3782	0.3134	0.2562	0.2068	0.1649
7	0.8095	0.7440	0.6728	0.5987	0.5246	0.4530	0.3856	0.3239	0.2687
8	0.8944	0.8472	0.7916	0.7291	0.6620	0.5925	0.5231	0.4557	0.3918
9	0.9462	0.9161	0.8774	0.8305	0.7764	0.7166	0.6530	0.5874	0.5218
10	0.9747	0.9574	0.9332	0.9015	0.8622	0.8159	0.7634	0.7060	0.6453
11	0.9890	0.9799	0.9661	0.9466	0.9208	0.8881	0.8487	0.8030	0.7520
12	0.9955	0.9912	0.9840	0.9730	0.9573	0.9362	0.9091	0.8758	0.8364
13	0.9983	0.9964	0.9929	0.9872	0.9784	0.9658	0.9486	0.9261	0.8981
14	0.9994	0.9986	0.9970	0.9943	0.9897	0.9827	0.9726	0.9585	0.9400
15	0.9998	0.9995	0.9988	0.9976	0.9954	0.9918	0.9862	0.9780	0.9665
16	0.9999	0.9998	0.9996	0.9990	0.9980	0.9963	0.9934	0.9889	0.9823
17	1.0000	0.9999	0.9998	0.9996	0.9992	0.9984	0.9970	0.9947	0.9911
18		1.0000	0.9999	0.9999	0.9997	0.9994	0.9987	0.9976	0.9957
19			1.0000	1.0000	0.9999	0.9997	0.9995	0.9989	0.9980
20					1.0000	0.9999	0.9998	0.9996	0.9991
21						1.0000	0.9999	0.9998	0.9996
22							1.0000	0.9999	0.9999
23								1.0000	0.9999
24									1.0000

جدول III : مجموع الاحتمال البواسوني (تابع)

r	μ								
	10.0	11.0	12.0	13.0	14.0	15.0	16.0	17.0	18.0
0	0.0000	0.0000	0.0000						
1	0.0005	0.0002	0.0001	0.0000	0.0000				
2	0.0028	0.0012	0.0005	0.0002	0.0001	0.0000	0.0000		
3	0.0103	0.0049	0.0023	0.0010	0.0005	0.0002	0.0001	0.0000	0.0000
4	0.0293	0.0151	0.0076	0.0037	0.0018	0.0009	0.0004	0.0002	0.0001
5	0.0671	0.0375	0.0203	0.0107	0.0055	0.0028	0.0014	0.0007	0.0003
6	0.1301	0.0786	0.0458	0.0259	0.0142	0.0076	0.0040	0.0021	0.0010
7	0.2202	0.1432	0.0895	0.0540	0.0316	0.0180	0.0100	0.0054	0.0029
8	0.3328	0.2320	0.1550	0.0998	0.0621	0.0374	0.0220	0.0126	0.0071
9	0.4579	0.3405	0.2424	0.1658	0.1094	0.0699	0.0433	0.0261	0.0154
10	0.5830	0.4599	0.3472	0.2517	0.1757	0.1185	0.0774	0.0491	0.0304
11	0.6968	0.5793	0.4616	0.3532	0.2600	0.1848	0.1270	0.0847	0.0549
12	0.7916	0.6887	0.5760	0.4631	0.3585	0.2676	0.1931	0.1350	0.0917
13	0.8645	0.7813	0.6815	0.5730	0.4644	0.3632	0.2745	0.2009	0.1426
14	0.9165	0.8540	0.7720	0.6751	0.5704	0.4657	0.3675	0.2808	0.2081
15	0.9513	0.9074	0.8444	0.7636	0.6694	0.5681	0.4667	0.3715	0.2867
16	0.9730	0.9441	0.8987	0.8355	0.7559	0.6641	0.5660	0.4677	0.3750
17	0.9857	0.9678	0.9370	0.8905	0.8272	0.7489	0.6593	0.5640	0.4686
18	0.9928	0.9823	0.9626	0.9302	0.8826	0.8195	0.7423	0.6550	0.5622
19	0.9965	0.9907	0.9787	0.9573	0.9235	0.8752	0.8122	0.7363	0.6509
20	0.9984	0.9953	0.9884	0.9750	0.9521	0.9170	0.8682	0.8055	0.7307
21	0.9993	0.9977	0.9939	0.9859	0.9712	0.9469	0.9108	0.8615	0.7991
22	0.9997	0.9990	0.9970	0.9924	0.9833	0.9673	0.9418	0.9047	0.8551
23	0.9999	0.9995	0.9985	0.9960	0.9907	0.9805	0.9633	0.9367	0.8989
24	1.0000	0.9998	0.9993	0.9980	0.9950	0.9888	0.9777	0.9594	0.9317
25		0.9999	0.9997	0.9990	0.9974	0.9938	0.9869	0.9748	0.9554
26		1.0000	0.9999	0.9995	0.9987	0.9967	0.9925	0.9848	0.9718
27			0.9999	0.9998	0.9994	0.9983	0.9959	0.9912	0.9827
28			1.0000	0.9999	0.9997	0.9991	0.9978	0.9950	0.9897
29				1.0000	0.9999	0.9996	0.9989	0.9973	0.9941
30					0.9999	0.9998	0.9994	0.9986	0.9967
31					1.0000	0.9999	0.9997	0.9993	0.9982
32						1.0000	0.9999	0.9996	0.9990
33							0.9999	0.9998	0.9995
34							1.0000	0.9999	0.9998
35								1.0000	0.9999
36									0.9999
37									1.0000

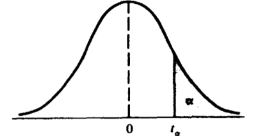


جدول IV : المساحة تحت المنحنى الطبيعي

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
-3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
-3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
-3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
-3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
-2.9	0.0019	0.0018	0.0017	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
-2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
-2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
-2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
-2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
-2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
-2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
-2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
-2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
-2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
-1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
-1.8	0.0359	0.0352	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
-1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
-1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
-1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
-1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0722	0.0708	0.0694	0.0681
-1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
-1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
-1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
-1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
-0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
-0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
-0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
-0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
-0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
-0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
-0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
-0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
-0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
-0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9278	0.9292	0.9306	0.9319

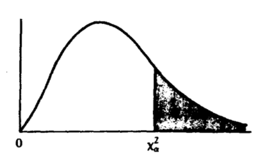
جدول IV : المساحة تحت المنحنى الطبيعي (تابع)

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998



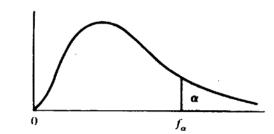
جدول ٧ : القيم الحرجة في توزيع t

ν	α 0.10	0.05	0.025	0.01	0.005
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
inf.	1.282	1.645	1.960	2.326	2.576



جدول VI : القيم الحرجة في توزيع كاي مربع

ν	α = 0.995	0.99	0.975	0.95	0.05	0.025	0.01	0.005
1	0.0^4393	0.0^3157	0.0^3982	0.0^2393	3.841	5.024	6.635	7.879
2	0.0100	0.0201	0.0506	0.103	5.991	7.378	9.210	10.597
3	0.0717	0.115	0.216	0.352	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	11.070	12.832	15.086	16.750
6	0.676	0.872	1.237	1.635	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	16.919	19.023	21.666	23.[illegible]
10	2.156	2.558	3.247	3.940	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	33.924	36.781	40.289	42.796
23	9.260	10.196	11.689	13.091	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	36.415	39.364	42.980	45.558
25	10.520	11.524	13.120	14.611	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	38.885	41.923	45.642	48.290
27	11.808	12.879	14.573	16.151	40.113	43.194	46.963	49.645
28	12.461	13.565	15.308	16.928	41.337	44.461	48.278	50.993
29	13.121	14.256	16.047	17.708	42.557	45.722	49.588	52.336
30	13.787	14.953	16.791	18.493	43.773	46.979	50.892	53.672



جدول VII : القيم الحرجة في توزيع f

$f_{0.05}(\nu_1, \nu_2)$

ν_2	ν_1 = 1	2	3	4	5	6	7	8	9
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88

جدول VII : القيم الحرجة في توزيع f (تابع)

$f_{0.05}(\nu_1, \nu_2)$

ν_2	ν_1 = 10	12	15	20	24	30	40	60	120	∞
1	241.9	243.9	245.9	248.0	249.1	250.1	251.1	252.2	253.3	254.3
2	19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50
3	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.36
6	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.2[illegible]
8	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.9[illegible]
9	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
26	2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69
27	2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67
28	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65
29	2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64
30	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120	1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25
∞	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

جدول VII : القيم الحرجة في توزيع f (تابع)

$f_{0.01}(\nu_1, \nu_2)$

ν_2	ν_1 = 1	2	3	4	5	6	7	8	9
1	4052	4999.5	5403	5625	5764	5859	5928	5981	6022
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41

جدول VII : القيم الحرجة في توزيع f (تابع)

$f_{0.01}(\nu_1, \nu_2)$

ν_2	ν_1									
	10	12	15	20	24	30	40	60	120	∞
1	6056	6106	6157	6209	6235	6261	6287	6313	6339	6366
2	99.40	99.42	99.43	99.45	99.46	99.47	99.47	99.48	99.49	99.50
3	27.23	27.05	26.87	26.69	26.60	26.50	26.41	26.32	26.22	26.13
4	14.55	14.37	14.20	14.02	13.93	13.84	13.75	13.65	13.56	13.46
5	10.05	9.89	9.72	9.55	9.47	9.38	9.29	9.20	9.11	9.02
6	7.87	7.72	7.56	7.40	7.31	7.23	7.14	7.06	6.97	6.88
7	6.62	6.47	6.31	6.16	6.07	5.99	5.91	5.82	5.74	5.65
8	5.81	5.67	5.52	5.36	5.28	5.20	5.12	5.03	4.95	4.86
9	5.26	5.11	4.96	4.81	4.73	4.65	4.57	4.48	4.40	4.31
10	4.85	4.71	4.56	4.41	4.33	4.25	4.17	4.08	4.00	3.91
11	4.54	4.40	4.25	4.10	4.02	3.94	3.86	3.78	3.69	3.60
12	4.30	4.16	4.01	3.86	3.78	3.70	3.62	3.54	3.45	3.36
13	4.10	3.96	3.82	3.66	3.59	3.51	3.43	3.34	3.25	3.17
14	3.94	3.80	3.66	3.51	3.43	3.35	3.27	3.18	3.09	3.00
15	3.80	3.67	3.52	3.37	3.29	3.21	3.13	3.05	2.96	2.87
16	3.69	3.55	3.41	3.26	3.18	3.10	3.02	2.93	2.84	2.75
17	3.59	3.46	3.31	3.16	3.08	3.00	2.92	2.83	2.75	2.65
18	3.51	3.37	3.23	3.08	3.00	2.92	2.84	2.75	2.66	2.57
19	3.43	3.30	3.15	3.00	2.92	2.84	2.76	2.67	2.58	2.49
20	3.37	3.23	3.09	2.94	2.86	2.78	2.69	2.61	2.52	2.42
21	3.31	3.17	3.03	2.88	2.80	2.72	2.64	2.55	2.46	2.36
22	3.26	3.12	2.98	2.83	2.75	2.67	2.58	2.50	2.40	2.31
23	3.21	3.07	2.93	2.78	2.70	2.62	2.54	2.45	2.35	2.26
24	3.17	3.03	2.89	2.74	2.66	2.58	2.49	2.40	2.31	2.21
25	3.13	2.99	2.85	2.70	2.62	2.54	2.45	2.36	2.27	2.17
26	3.09	2.96	2.81	2.66	2.58	2.50	2.42	2.33	2.23	2.13
27	3.06	2.93	2.78	2.63	2.55	2.47	2.38	2.29	2.20	2.10
28	3.03	2.90	2.75	2.60	2.52	2.44	2.35	2.26	2.17	2.06
29	3.00	2.87	2.73	2.57	2.49	2.41	2.33	2.23	2.14	2.03
30	2.98	2.84	2.70	2.55	2.47	2.39	2.30	2.21	2.11	2.01
40	2.80	2.66	2.52	2.37	2.29	2.20	2.11	2.02	1.92	1.80
60	2.63	2.50	2.35	2.20	2.12	2.03	1.94	1.84	1.73	1.60
120	2.47	2.34	2.19	2.03	1.95	1.86	1.76	1.66	1.53	1.38
∞	2.32	2.18	2.04	1.88	1.79	1.70	1.59	1.47	1.32	1.00

جدول VIII : عوامل التحمل في التوزيع الطبيعي

$\nu = 0.95$

n	$1-\alpha$ 0.90	0.95	0.99
2	32.019	37.674	48.430
3	8.380	9.916	12.861
4	5.369	6.370	8.299
5	4.275	5.079	6.634
6	3.712	4.414	5.775
7	3.369	4.007	5.248
8	3.136	3.732	4.891
9	2.967	3.532	4.631
10	2.839	3.379	4.433
11	2.737	3.259	4.277
12	2.655	3.162	4.150
13	2.587	3.081	4.044
14	2.529	3.012	3.955
15	2.480	2.954	3.878
16	2.437	2.903	3.812
17	2.400	2.858	3.754
18	2.366	2.819	3.702
19	2.337	2.784	3.656
20	2.310	2.752	3.615
25	2.208	2.631	3.457
30	2.140	2.549	3.350
35	2.090	2.490	3.272
40	2.052	2.445	3.213

$\nu = 0.99$

n	$1-\alpha$ 0.90	0.95	0.99
2	160.193	188.491	242.300
3	18.930	22.401	29.055
4	9.398	11.150	14.527
5	6.612	7.855	10.260
6	5.337	6.345	8.301
7	4.613	5.488	7.187
8	4.147	4.936	6.468
9	3.822	4.550	5.966
10	3.582	4.265	5.594
11	3.397	4.045	5.308
12	3.250	3.870	5.079
13	3.130	3.727	4.893
14	3.029	3.608	4.737
15	2.945	3.507	4.605
16	2.872	3.421	4.492
17	2.808	3.345	4.393
18	2.753	3.279	4.307
19	2.703	3.221	4.230
20	2.659	3.168	4.161
25	2.494	2.972	3.904
30	2.385	2.841	3.733
35	2.306	2.748	3.611
40	2.247	2.677	3.518

جدول VIII : عوامل التحمل في التوزيع الطبيعي (تابع)

$\gamma = 0.95$

n	$1-\alpha$		
	0.90	0.95	0.99
45	2.021	2.408	3.165
50	1.996	2.379	3.126
55	1.976	2.354	3.094
60	1.958	2.333	3.066
65	1.943	2.315	3.042
70	1.929	2.299	3.021
75	1.917	2.285	3.002
80	1.907	2.272	2.986
85	1.897	2.261	2.971
90	1.889	2.251	2.958
95	1.881	2.241	2.945
100	1.874	2.233	2.934
150	1.825	2.175	2.859
200	1.798	2.143	2.816
250	1.780	2.121	2.788
300	1.767	2.106	2.767
400	1.749	2.084	2.739
500	1.737	2.070	2.721
600	1.729	2.060	2.707
700	1.722	2.052	2.697
800	1.717	2.046	2.688
900	1.712	2.040	2.682
1000	1.709	2.036	2.676
∞	1.645	1.960	2.576

$\gamma = 0.99$

n	$1-\alpha$		
	0.90	0.95	0.99
45	2.200	2.621	3.444
50	2.162	2.576	3.385
55	2.130	2.538	3.335
60	2.103	2.506	3.293
65	2.080	2.478	3.257
70	2.060	2.454	3.225
75	2.042	2.433	3.197
80	2.026	2.414	3.173
85	2.012	2.397	3.150
90	1.999	2.382	3.130
95	1.987	2.368	3.112
100	1.977	2.355	3.096
150	1.905	2.270	2.983
200	1.865	2.222	2.921
250	1.839	2.191	2.880
300	1.820	2.169	2.850
400	1.794	2.138	2.809
500	1.777	2.117	2.783
600	1.764	2.102	2.763
700	1.755	2.091	2.748
800	1.747	2.082	2.736
900	1.741	2.075	2.726
1000	1.736	2.068	2.718
∞	1.645	1.960	2.576

ثبت المصطلحات

■ عربي / إنجليزي ■ إنجليزي / عربي .



عربي/إنجليزي

أ

احتمال	Probability
احتمال شرطي	Conditional probability
إحصاء	Statistics
الإحصاء الاستنتاجي	Deductive statistics
إحصاء وصفي	Desciptive statistics
اختبار ثنائي الجانب	Two-sided test
اختبار ثنائي الذيل	Two-tailed test
اختبار وحيد الجانب	One-sided test
اختبار وحيد الذيل	One-tailed test
الاختلاف	Variation
اختيار	Selection
أساس	Base
الانتشار	Dispersion
انحناء	Bending
الانحراف المعياري	Standard deviation

ت

تباديل	Permutations
تباين	Variance
تجزئة	Partition
تجزئيات مرتبة	Orderd partition
تحليل	Analysis
الترجع	Regression
الترقيم . العد	Counting
تقاطع	Intersection
تقدير	Estimation
تقدير غير متحيز	Unbiased estimator
تقدير متحيز	Biased estimator
تقريب	Approximation
تكرار	Repetition
التكرار النسبي	Relative frequency
توزيع احتمالي	Probability distribution
توزيع احتمالي مستمر	Continuous probability
توزيع احتمالي مشترك	Joint probability distribution
توزيع برنولي	Bernoulli distribution
توزيع تجريبي	Empericall distribution
توزيع تراكمي	Cumulative distribution
توزيع سالب	Negative distribution
توزيع شرطي	Conditional distribution
التوزيع العيني للوسط	Sampling distribution of mean
التوزيع العيني للنسبة	Sampling distribution of proportion
توزيع طبيعي	Normal distribution
توزيع غما	Gamma distribution
توزيع غوص	Gaussian distribution
توزيع كاي – مربع	Chi-Square distribution
توزيع المعاينة	Sampling
توزيع المعاينة التجريبي	Exprimental sampling
توزيع هامشي	Marginal distribution
توزيع هندسي زائدي	Hypergeometric
التوقع	Expectation

ث

ثابت Constant

ج

جداء (ضرب) Product

ح

حادث Event

حادث أولي Elementary event

حادث مستحيل Impossible event

حدّاني Binomial

حوادث متنافية تبادلياً Mutually exclusive events

حوادث مستقلة Independent events

خ

خط تنازلي Asymptote

د

دالة احتمالية Probability function

دالة تكرارية Frequencey function

دالة جبرية Algebraic function

دالة خطية Linear function

دالة درجية Step function

دالة الكثافة Density function

دالة الكثافة الاحتمالية Probability density function

دالة مستمرة Continuous function

درجات من الحرية Degrees of freedom

ذ

الذيل (وجه الكتابة في قطعة النقود) Tail

ر

الرأس (وجه النقش في قطعة النقود) Head

ز

زاوية Angle

س

سالب Negative

السرعة الزاوية Angular velocity

السحب بدون الإعادة Sampling without replacement

السحب مع الإعادة Sampling with replacement

سلسلة Chain

ش

شعاع Vector

شكل بياني Graph

شيء – هدف Object

ض

ضرب (جداء) Product

ع

العد . الترقيم Counting

عزم Moment

عزم حول الصفر Moment about zero

عزم حول وسط Moment about mean

عمود Column

عضو Member

عنصر Component

عينة Sample

عينة عشوائية Random sample

غ

غير قابل للعد Uncountable

غير محدود (لانهائي) Infinite

ف

الفرض البديل	Alternative hypothesis
فرض العدم	Null hypothesis
فرضيات	Hypotheses
الفرضيات الاحصائية	Statistical hypotheses
فضاء العينة	Sample space
فك حداني	Binomial expansion

ق

قابل للعد	Countable
القرارات الإحصائية	Statistical decision
قوة	Power
قيمه حرجة	Critical number
قيمة متوقعة	Expected value

م

متباينة	Inequality
متتالية	Sequence
متغيرات مستقلة	Independent variables
متغير عشوائي	Random variable
متغير عشوائي منقطع	Discrete random variable
متغير مرتبط	Dependent variable
مجال ثقة	Confidence interval
مجتمع إحصائي	Statistical population
مجموعة	Set
مجموعة خالية	Null set
محدود	Finite
محوّل	Transformed
مخطط فين	Venn diagram
مستقل	Independent
مستمر	Continuous
مصفوفة	Matrix
معيب	Defective
ملاحظة	Observation
مَتَاسب الحجم	Volume prelatives
منطقة حرجة	Critical region
منطقة رفض الفرضية	Region of rejection of hypothesis

ن

نظرية القرارات	Decisions Theory

هـ

هندسة تحليلية	Analytic Geometry

و

وحيد المنوال	Unimodal
وسط توافقي	Harmonic mean
وسيط	Median

ى

يتقارب	Converge

sampling distribution of proportion — التوزيع العينى للنسبة
sample space — فضاء العينة
sampling with replacement — السحب مع إعادة
sampling without replacement — السحب بدون إعادة
selection — اختبار
set — مجموعة – فئة
sequnce — متتالية
subset — مجموعة جزئية – فئة جزئية
slope — ميل
standard — معيارى
standard deviation — الانحراف المعيارى
standarized random variable — متغير عشوائى مّعيارى
stationary — مستقر
statistics — إحصاء
statistical decisions — القرارات الإحصائية
statistical hypotheses — الفرضيات الإحصائية
stochastic process — عملية عشوائية
step function — دالة درجية

T

tail — الذيل – وجه الكتابة فى قطعة النقود
test — اختبار
test of hypotheses — اختبار الفرضيات
test of significance — اختبار المعنوية
test statistic — إحصاء الاختبار
theory of decisions — نظرية القرارات
transformed — محول
two-tailed test — اختبار ثنائى الذيل
two-sided test — اختبار ثنائى الجانب

U

unbiased estimator — تقدير غير متحيز
uncorrelated — غير مرتبط
uncountable — غير قابل للعد
unimodal — وحيد المنوال
union — اجتماع – اتحاد
universe — المجموعة الكلية

V

value — قيمة
variable — متغير
variance — تباين
variation — الاختلاف
vector — شعاع
Venndiagram — مخطط فين
volume relatives — مناسب الحجم

W

weight — وزن

X

x-co-ordinate — الاحداثى السينى

Y

y-cordinate — الاحداثى الصادى

Z

z-co-ordinate — الاحداثى العينى
zero order correlation coefficients — معاملات الارتباط من الرتبة صفر
zero point — نقطة الصفر
zero vector — شعاع صفرى

measurements قياسات
median الوسيط
member عضو
moment العزم
moment about the mean العزم حول الوسط
moment about the zero العزم حول الصفر
multinomial كثير حدود
mutually exclusive events حوادث متنافية تبادليا

N

negative سالب
negative distribution التوزيع السالب
nondiscrete غير منقطع
normal distribution التوزيع الطبيعى
null hypotheses فرض العدم
null set المجموعة الخالية
number of degrees of freedom عدد درجات الحرية

O

object شيء – هدف
observation ملاحظة
one-tailed test اختبار وحيد الذيل
one-sided test اختبار وحيد الجانب
operation عملية
ordered partitions تجزئيات مرتبة
ordinate الاحداثى الصادى
outcome نتيجة

P

partition تجزئة
perfect تام
permutations تباديل
population مجتمع إحصائى
probability الاحتمال
probability density function دالة الكثافة الاحتمالية
probability distribution التوزيع الاحتمالى
probability function دالة احتمالية
product جداء – ضرب
power قوة

Q

quadratic mean الوسط التربيعى

R

random variable متغير عشوائى
random sample عينة عشوائية
range مدى
real number عدد حقيقى
rectangle مستطيل
rectangular-coordinates الإحداثيات المتعامدة
region منطقة
region of rejection of the hypothesis منطقة رفض الفرضية
regression الترجيع
relative frequency التكرار النسبى
reqeated tests اختبارات متتالية
repetition تكرار

S

sample عينة
sampling distribution توزيع المعاينة
sampling distribution of mean التوزيع العينى للوسط

density function	دالة الكثافة
dependent variable	متغير مرتبط
descriptive statistics	الإحصاء الوصفى
diagram	شكل بيانى
difference	فرق
discrete random variable	متغير عشوائى منقطع
dispersion	الانتشار
distribution	التوزيع

E

elementary event	حادث أولى
empirically distribution	التوزيع التجريبى
estimation	تقدير
even	زوجى
event	حادث
expectation	التوقع
expected value	القيمة المتوقعة
experimental sampling distribution	توزيع المعاينة التجريبى

F

factorial	عاملى
finite	محدود
frequency	تكرارى
frequency function	دالة تكرارية

G

games of chance	ألعاب الحظ
gamma distribution	توزيع غما
Gaussian distrbution	توزيع غوص
geometrical measurement	مقياس هندسى
graph	شكل بيانى

H

harmonic mean	الوسط التوافقى
head	الرأس (وجه النقش في قطعة النقود)
hypothesis	فرضيات
hypergeometric distribution	التوزيع الهندسى الزائدى

I

identity	وحدة
impossible event	حادث مستحيل
independent	مستقل
independent events	حوادث مستقلة
independent variables	متغيرات مستقلة
inequality	متباينة
infinite	غير محدود (لا نهائى)
initial	أولى
intersection	تقاطع
item	عنصر ــ بند ــ فقرة

J

joint probobility distribution	التوزيع الاحتمالى المشترك

L

line graph	خط بيانى
linear function	دالة خطية
list	قائمة

M

main	رئيسى
main diagonal	القطر الرئيسى
marginal distribution	التوزيع الهامشى
matrix	مصفوفة

إنجليزي/عربي

A

abscissa	الاحداثي السيني
absolute value	القيمة المطلقة
algebra	الجبر
algebraic function	دالة جبرية
alternative hypothesis	الفرض البديل
angle	زاوية
analytic geometry	هندسة تحليلية
analysis	تحليل
angular velocity	السرعة الزاوية
approximation	تقريب
asymtote	خط مقارب

B

biased estimator	تقدير متحيز
base	أساس
belong	ينتمي
bending	انحناء
Bernoulli distribution	توزيع برنولي
binomial	حداني
binomial coefficients	معاملات حدانية
binomial distribution	توزيع حداني
binomial expansion	مفكوك حداني

C

central limit theorem	نظرية النهاية المركزية
chain	سلسلة
characterisitc function	دالة مميزة
chi-square distribution	توزيع كاي — مربع
coefficient	معامل
closed interval	مجال مغلق
column	عمود
component	عنصر
conditional distribution	التوزيع الشرطي
conditional probability	الاحتمال الشرطي
confidence interval	مجال ثقة
confidence level	مستوى ثقة
constant	ثابت
continuous	مستمر
continuous function	دالة مستمرة
continuous probability distribution	التوزيع الاحتمالي المستمر
continuous variable	متغير مستمر
converge	يتقارب
countable	قابل للعد
count	يعد — يحصي
counting	العد — الترقيم
critical region	منطقة حرجة
critical value	قيمة حرجة
cumulative distribution	توزيع تراكمي

D

deductive statistics	الإحصاء الاستنتاجي
defective	معيب





مطابع جامعة الملك عبد العزيز





مطابع جامعة الملك عبد العزيز